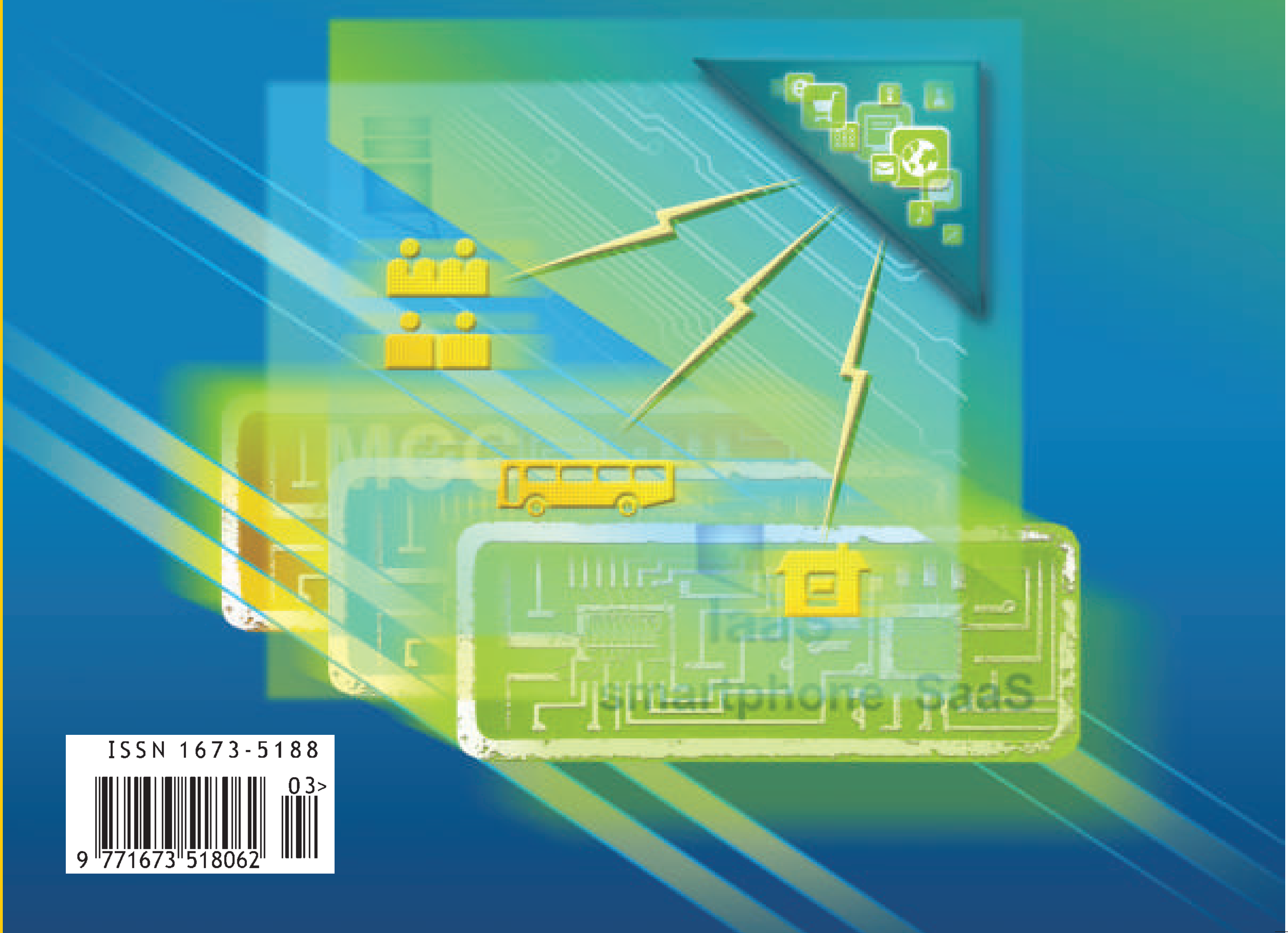
www.zte.com.cn/magazine/English

ISSN 1673-5188
CODEN ZCTOAK

ZTE COMMUNICATIONS

March 2011, Vol.9 **No.1**

Special Topic: Mobile Cloud Computing and Applications



ISSN 1673-5188
03>
9 771673 518062

ZTE中兴
ZTE COMMUNICATIONS
VOLUME 9 NUMBER 1 MARCH 2011

ZTE Ranked Second Worldwide for International Patent Applications

ZTE announced on February 11 that, according to the World Intellectual Property Organization (WIPO), it had leapt from 23rd place in 2009 to second place in 2011 in terms of the number of international patent applications submitted. This follows a significant jump in the overall number of international patent applications from China, now the world's fourth largest submitter of patent applications.

As of December 31, 2010, ZTE had 1,863 international patent applications registered with WIPO.

According to official data from WIPO, the number of Patent Cooperation Treaty (PCT) applications submitted by ZTE has leapfrogged for five years in a row. Since 2006, when ZTE's PCT ranking entered the top 100 for the first time, its annual ranking has jumped from 52nd in 2007 to second place in 2010. ZTE's global competitiveness has also been growing proportionally, with the percentage of overseas revenue increasing from 44% in 2006 to 57.8% in 2007and 60% in 2008.

Despite the global economic slowdown over the past two years, ZTE has still realized steady growth—global revenue from principal operations reached RMB 70.3 billion in 2010 and has increased 16.69% year-on-year.

ZTE has a standard practice of investing 10% of its income into R&D, including the development of its international patent strategy. Even during the financial crisis, ZTE did not reduce R&D spending. Total R&D spending in these two years reached RMB 13 billion. WIPO noted that due to the financial crisis and a decline in R&D spending, European and North American countries did not perform well in terms of patent applications. The number of international patent applications from the U.S. and UK has been in decline for three consecutive years. In comparison, the number of international patent applications from China has grown by 30% and 56% over the past two years respectively. This is faster than any other country in the world. (ZTE Corporation)

ZTE Ranked Third Globally Among LTE Infrastructure Vendors

ZTE Corporation announced on February 18 that it is now among the top three vendors in the LTE infrastructure industry, according to a recently announced market tracker report by Open Vista Consulting.

According to the report, ZTE has achieved outstanding results in industry contribution, market competitiveness, product competitiveness, and solution competitiveness. These results show that ZTE has moved to become one of the top three LTE infrastructure vendors in the world.

ZTE accounts for 21% of the LTE market and "has made considerable contributions to TD-LTE in the evolution from TD to LTE," the report states. It also states that ZTE was the first to put forward and commercially deploy Software Defined Radio (SDR) and that ZTE has enabled voice support in LTE by applying Evolved Packet Core (EPC) and IP Multimedia Subsystem (IMS).

The report also states ZTE has one of the most complete product lineups of any vendor and therefore is in a leading position in terms of platform products and TCO.

ZTE envisions that LTE will be a strategic product in the mobile broadband era. The company has 4,000 researchers involved in R&D and global promotion of LTE products. Building LTE test networks and commercial networks throughout the world is part of the company's LTE strategy. In 2011, it will further promote the application of LTE around the world. (ZTE Corporation)

ZTE Recognized for Industry Best Practices as Frost & Sullivan's SDR Equipment Vendor of the Year

ZTE Corporation announced on February 9 that it had received the Best Practice Award for being Frost & Sullivan's "2010 SDR Equipment Vendor of the Year." This award recognizes ZTE's outstanding contribution to the development of Software Defined Radio (SDR) technology and the wireless communications industry.

At the end of 2010, shipments of ZTE SDR base stations exceeded 700,000 units, triple that of 2009. ZTE was the first in the world to introduce the SDR base station platform and also the first to realize its commercial application. The SDR base station platform supports GSM, UMTS, CDMA2000, TD-SCDMA, FDD LTE, WiMAX, and TDD LTE wireless access modes on a single hardware platform.

In recent years, ZTE has achieved comprehensive breakthroughs in the wireless communications market by leveraging the cutting-edge capabilities of this new platform. In Europe, ZTE has mass-delivered SDR base stations to KPN, Telenor, Telefonica, Optimus, and H3G. In China, ZTE is ranked first in new UMTS market share and second in total market share.

ZTE continues to be the largest shipper of CDMA products. Its WiMAX products have been adopted by mainstream operators such as KDDI, Clearwire, and Telefonica. ZTE's LTE products have also been accepted for commercial application and test networks in 12 countries.

ZTE's SDR base station is now deployed in Germany and is used by more than 60 operators in countries such as Belgium, Spain, Portugal, Poland, Turkey, Japan, Russia, Canada, and the U.S.

The SDR vendor ranking is maintained by Frost & Sullivan, a consulting firm with more than 35 offices in six continents. Frost and Sullivan provides marketing and strategic consulting services for global 1000 companies, startups, and investment organizations. (ZTE Corporation)

ZTE Exclusively Demonstrates LTE-A CoMP-Based Services at Mobile World Congress 2011

ZTE Corporation demonstrated video call, video streaming, and ftp services based on LTE-A Coordinated Multipoint (CoMP) technology at "Mobile World Congress 2011" in Barcelona, Spain. This makes ZTE the first and only telecoms vendor in the world to demonstrate LTE-A CoMP-based services.

During the demonstration, the system achieved downstream peak speed of over 1 Gbit/s, and participants were the first to experience the lightning fast speeds of fourth generation wireless technologies. ZTE also announced plans to introduce commercial LTE-A products in 2012 in order to meet the growing need for high-speed data communication.

ZTE's LTE-A CoMP prototype incorporates a number of key technologies including multiantenna enhancement, MU-MIMO, and multipoint joint coordinated transmission. By using CoMP technology, throughput at the edge is dramatically increased.

In recent years, ZTE has accumulated technologies in core LTE areas. As of November 30, 2010, ZTE held 235 ETSI Essential Patents (EPs), accounting for 7% of the total number of EPs for LTE that had already been declared. This shows that the corpoartion is a first tier holder of EPs for 4G standards. To date, ZTE has contributed about 550 proposals to the 3GPP RAN Working Group in LTE-A. (ZTE Corporation)

Contents

Http://www.zte.com.cn/magazine/English
Email: magazine@zte.com.cn

ZTE COMMUNICATIONS
Vol. 9 No.1 (Issue 29)
Quarterly
First Issue Published in 2003

Supervised by:
Anhui Science and Technology Department

Sponsored by:
ZTE Corporation and Anhui Science and Technology Information Research Institute

Staff Members:
Editor-in-chief: Xie Daxiong
Deputy Editor-in-chief: Deng Xin
Executive Deputy Editor-in-chief: Huang Xinming
Editor in Charge: Zhu Li
Editors: Paul Sleswick, Yang Qinyi, Xu Ye, Lu Dan
Producer: Yu Gang
Circulation Executive: Wang Pingping
Assistant: Wang Kun

Editorial Correspondence:
Add: 450 Rongshida Avenue, Hefei 230041, P. R. China
Tel: +86-551-5533356
Fax: +86-551-5850139
Email: magazine@zte.com.cn

Published and Circulated (Home and Abroad) by:
Editorial Office of ZTE COMMUNICATIONS
Printed by:
Hefei Zhongjian Color Printing Company
Publication Date: March 25, 2011
Publication Licenses:
ISSN 1673-5188
CN 34-1294/TN
Advertising License:
皖合工商广字0058号
Annual Subscription Rate:
USD$50

GUEST PAPER

Special Topic: Mobile Cloud Computing and Applications

Research Papers

Development Field

Operational Application

Lecture Series

Roundup

Departments

Editor's Desk: Tri-network convergence is becoming a hot topic as interest in 3G, the Internet of things, and cloud computing continues to grow. Most countries now place great emphasis on tri-network convergence and one after the other have prioritized it as a means of developing new strategic sectors. With the arrival of 2011, our magazine has the honor of inviting Mr. Hequan Wu, a telecommunications expert from the Chinese Academy of Engineering to deliver a speech on the characteristics and challenges of tri-network convergence in China.

According to Mr. Wu, some developed countries have already achieved tri-network convergence and others are promoting it. China has its own characteristics in implementing tri-network convergence, and these characteristics are bringing new management modes and new challenges.

Mr. Wu pointed out that new management modes were needed in the process of tri-network convergence in China, and the introduction of the integrated network broadcasting control platform would bring about new challenges to management modes. Adaption to the platform requires cooperation and innovation.

Finally, Mr. Wu emphasized the key to the success of Chinese tri-network convergence was to develop new services and business models.

Chinese Tri-Network Convergence: Characteristics and Challenges

Hequan Wu

Hequan Wu is a member of the Chinese Academy of Engineering (CAE). He researches the development of digital and optical transmission systems and in recent years has presided over next-generation Internet, 3G, and LTE projects. He was the former vice-president of the CAE and is now a deputy director of the China Advisory Committee for State Informatization, a consultant of the Communications Science & Technology Commission of the Ministry of Industry and Information Technology of China, a deputy director of the Chinese Institute of Communications and the Chinese Institute of Electronics, director of China Communications Standards Association, a leader of Tri-Network convergence Expert Group, and director of China Next Generation Internet (CNGI) Experts Committee.

In the *2010 Report on the Work of the Government*, delivered by Wen Jiabao, tri-network convergence is interpreted as "the integration of telecommunications networks, cable television networks, and the Internet" [1]. Enabling these three networks to interconnect with each other and share resources will provide users with diverse voice, data, radio, and TV services. Pilot projects have already been launched in China with the aims of integrating radio, TV services, and telecommunication services; quickening overall planning and reconstruction of the networks, and reinforcing monitoring of information and cultural security. The ultimate goal is to promote the development of related sectors.

Some developed countries have already achieved tri-network convergence, and others are promoting it. Tri-network convergence will bring about new management modes and new challenges, and China has its own characteristics in implementing it. First, only a small number of broadcast and TV networks are digital and connected to telecom networks. These are prerequisite for broadcast and TV networks to deliver telecom services. According to the State Administration of Radio, Film and TV, Next Generation Broadcasting (NGB) networks will be constructed to connect radio and TV networks. NGB networks adopt IP technologies to support telecom services. However, adopting packet technologies in broadcast video transmission is untried, and its cost-effectiveness is yet to be assessed.

Second, the Chinese government pays special attention to security. Transforming a TV set into an Internet terminal opens the possibility that it may be attacked by hackers or viruses. Undesirable or even illegal content, which cannot be completely eliminated in the open Internet, may appear on the television screen at home. However, the existing security platform, technical infrastructure, and guaranteed mechanism of broadcast and TV networks are not capable of processing a sharp increase in video flow that would arise as a result of network convergence. The existing broadcasting control mechanism also needs to be improved to meet the demands of operating multiple services. In short, both information security and network security issues have to be addressed.

Third, a new management mode is required for tri-network convergence. For content security, Chinese authorities have issued regulations stipulating that for a telecom operator to transmit broadcast and TV services, its video services must be managed by the integrated broadcasting control platform of the Broadcast and TV Authority. However, some issues are not clearly defined and need further clarification; for example, to what extent is content controlled and what interfaces and functions should the platform provide? The benefits of IPTV lie in the Internet-based value-added services that come with it. But these services may not be video services and need not be managed and operated by the platform. The problem arises of how to coordinate value-added services with IPTV. According to the requirements of the pilot projects, the integrated broadcasting control platform should manage not only content, but also users and billing. This means a user may be managed by both the telecom operator and the broadcast and TV operator when accessing a service. This is a great challenge to the network management system and client management mechanism. Therefore, it is necessary to coordinate management and services of the broadcast and TV operator with those of the telecom operator. In short, an integrated broadcasting control platform brings new challenges to the management mode. Meeting these challenges requires cooperation and innovation. It requires reform of management mechanisms of administrative bodies and reform of the technical and service modes of operators.

Fourth, the pilot projects have clear requirements in terms of access rate: In 2012, the downlink and broadband access rates for broadcast and TV services will exceed 2 Gbit/s and 100 Mbit/s respectively; while the downlink and household access rates for telecom services will reach 1 Gbit/s and 100 Mbit/s respectively. As Chinese urban areas are densely populated, the costs of Fiber to the Building (FTTB) and Fiber to the Zone (FTTZ) will be much lower than those in developed countries. FTTB and FTTZ have become the main access methods in tri-network convergence programs in China. A telecom operator may adopt FTTB or FTTZ plus Digital Subscriber Line (DSL) or Local Area Network (LAN), while a broadcast and TV operator may use FTTB or FTTZ plus Ethernet over Coax (EoC). For FTTB/FTTZ, Ethernet Passive Optical Network (EPON) or Gigabit Passive Optical Network (GPON) or a combination of these is suitable for densely populated areas in China. Because tri-network convergence involves a huge number of users, developing low-cost access systems such as PON is key for success. The household access rate of 100 Mbit/s is not difficult to achieve in an optical access system, but it is a great challenge for core networks, especially Metropolitan Area Networks (MANs). The rate requirement of 100 Mbit/s is higher than that in some developed countries, and high bandwidth requires high value and high return on investment. Therefore, we should explore advanced technologies that are suitable for China's situation and develop new broadband services and business models that bring value to users and provide sustainable return to operators. According to a statistics report of the China Internet Network Information Center (CNNIC), Chinese netizens spend around 4% of their disposable income on the Internet. If mobile communications are included, the ratio of communications expenses to income is higher than in some developed countries. Despite this, telecom operators are suffering from a lack of profit-making modes. Therefore, the development of new services and business models is key to the success of tri-network convergence.

Fifth, the multicasting problem needs to be solved. Other countries also encounter this problem when deploying IPTV services. But because of the huge number of users in China, existing telecom networks have to be reconstructed as manageable IP networks that can support large-scale multicasting. The network layer multicasting solution based on routers is complicated to implement. And while application layer multicasting based on terminals is easy to deploy and does not involve changing infrastructure, it does not perform as well as network layer multicasting in terms of stability and efficiency. Multicasting suffers problems such as multicast video source fraud and registration packet fraud by illegal users. In sum, it is necessary to develop a large-capacity and manageable IPTV multicast solution.

Sixth, Chinese telecom operators and broadcast and TV operators differ from their counterparts in other countries in monitoring mechanisms and scale. These differences impact network technologies, services and maintenance management mechanisms to be adopted.

In China, tri-network convergence is regarded as an important program for deepening reform in the telecom industry, fostering the emergence of strategic industries, and bringing benefit to the public. It is expected to drive the development of the telecom industry, and broadcast and TV industry. Achieving tri-network convergence demands the efforts of all parties in the telecom, and broadcast and TV industries. Chinese tri-network convergence has a long way to go and 2011 will be a critical year.

Reference

[1] Wen Jiabao. (2010, March). Report on the Work of the Government. [Online]. Available: http://www.gov.cn/english/official/2010-03/15/content_1556124_8.htm

Mobile Cloud Computing and Applications

Chengzhong Xu is a professor of Electrical and Computer Engineering and the Director of Cloud and Internet Computing Laboratory and Sun's Center of Excellence in Open Source Computing and Applications at Wayne State University. His research interests include distributed and parallel computing and wireless embedded systems. He has published two books and more than 150 papers and chaired numerous international conferences and workshops in these areas. He received a Ph.D. degree from the University of Hong Kong in 1993.

In 2010, cloud computing gained momentum. Cloud computing is a model for real-time, on-demand, pay-for-use network access to a shared pool of configurable computing and storage resources. It has matured from a promising business concept to a working reality in both the private and public IT sectors. The U.S. government, for example, has requested all its agencies to evaluate cloud computing alternatives as part of their budget submissions for new IT investment.

In recent years we have also witnessed the rapid growth of mobile applications due to the increasing popularity of smartphones and ubiquity of wireless access. Cloud computing fuels innovation in mobile computing and opens new pathways between mobile devices (where an application is launched) and the infrastructure (where data is stored and processed). Because mobile devices have intrinsic storage, processing, and battery power constraints, mobile applications often hit a performance wall. Unlimited computing and storage resources offered by cloud computing can help break through this wall and turn the problem into a vast opportunity for the growth of mobile computing. According to the latest study from Juniper Research, the market for cloud-based mobile applications is expected to grow 88% annually and reach $9.5 billion by 2014.

To a typical mobile user, a mobile application driven by the cloud should look and feel just like any native mobile applications installed and run in their mobile device. There are already some well-known cloud-based mobile applications; for example, Google's Gmail for iPhone and Cisco's WebEx on iPad. These are largely run as Software-as-a-Service (SaaS), in which a cloud provider's applications are deployed and run in the cloud and can be accessed by users. In general, cloud computing goes beyond the SaaS model by offering computing and storage Infrastructure as a Service (IaaS) or application development Platform as a Service (PaaS). Each cloud service model has proved efficacious in desktop computing. However, the benefits of IaaS and PaaS in mobile cloud computing have not been fully exploited.

This special issue of *ZTE Communications* discusses related issues in mobile cloud computing. The purpose is to provide an overview of this cutting edge field and to describe its development, trends, challenges, and current practices. Papers have been included that cover a broad spectrum of interesting topics, including mobile cloud computing architectures, mobile search and data management, energy management and sustainability, privacy and security, mobile social networks, and novel cloud-assisted smartphone applications.

In the paper, "A Survey of Mobile Cloud Computing," Fan et al. classify mobile cloud computing systems. Two representative systems, Hyrax and Cloudlet, are discussed in detail. In their paper "Mirroring Smartphones for Good: A Feasibility Study," Zhao et al. propose a framework that keeps a mirror for each smartphone on a computing infrastructure in the telecom network. In this framework, some computational workload is offloaded from a smartphone to its mirror. They demonstrate the efficacy of the framework in data caching applications and antivirus scanning services.

"A Cloud-Based Virtualized Execution Environment for Mobile Applications," by Hung et al. presents a cloud-based virtualized execution environment framework for mobile applications, with a focus on schemes for migrating applications and synchronizing data between execution environments. Performance and power saving issues involved in application migration are also discussed. In "Building a Platform to Bridge Low End Mobile Phones and Cloud Computing Services," Tso et al. propose a Thumb-in-Cloud platform to break the performance wall in low-end mobile phones. The platform consists of virtual machines that are deployed in low-end phones for execution of mobile applications. It also consists of Thumb gateways that tailor cloud services by reformatting and compressing the service content to fit into the phone's profile.

Zhang et al. in "WiFace: A Secure Geosocial Networking System Using Wi-Fi Based Multihop MANET," present a geosocial networking system running on a Wi-Fi based multihop ad hoc network platform for personal mobile devices. The system allows users to access cloud services in environments with or without networking infrastructure or GPS modules. In "A Case for Cloud-Based Mobile Search," Gao et al. design an Internet search case for cloud-based mobile applications. Searches launched in a mobile device invoke a cloud-based search engine to fulfill the tasks. Key enabling technologies are discussed.

"An On-Demand Security Mechanism for Cloud-Based Telecommunications Services," by Lin et al. investigates the security issues in cloud computing and a security model is proposed based on a security domain division concept. This helps provide dynamic, on-demand, and differentiated protection for services.

I am grateful to the authors who submitted for this special issue and to the reviewers who spent their valuable time to provide constructive feedback. I hope that you find this special issue interesting and useful.

A Survey of Mobile Cloud Computing

Xiaopeng Fan [1], *Jiannong Cao* [2], *and Haixia Mao* [3]
(1. Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences;
2. Department of Computing, Hong Kong Polytechnic University;
3. Department of Land Surveying and Geo-informatics, Hong Kong Polytechnic University)

Abstract: Mobile Cloud Computing (MCC) is emerging as one of the most important branches of cloud computing. In this paper, MCC is defined as cloud computing extended by mobility, and a new ad-hoc infrastructure based on mobile devices. It provides mobile users with data storage and processing services on a cloud computing platform. Because mobile cloud computing is still in its infancy, we aim to clarify confusion that has arisen from different views. Existing works are reviewed, and an overview of recent advances in mobile cloud computing is provided. We investigate representative infrastructures of mobile cloud computing and analyze key components. Moreover, emerging MCC models and services are discussed, and challenging issues are identified that will need to be addressed in future work.
Keywords: mobile cloud computing; cloud computing

1 Introduction

Since the first computer was invented in 1946, people have dreamt of using computing resources as a utility—like water, electricity, gas, or the telephone. There has been a long-held perception that computing will one day be the fifth utility. Cloud computing [1]-[6] might be the most promising way of realizing this dream.

In recent research on cloud computing, one of the most important issues has been how to sweep up data and programs from desktop PCs [1] and install them in the compute cloud. Cloud computing refers to applications delivered as services over the Internet and also the hardware and systems software in the data centers that provide those services [2]. However, whether it is called cloud computing or on-demand computing, Software as a Service (SaaS) or Internet as platform, the overriding idea is a shift in the geography of computation.

In general, cloud computing is web-based processing [7], whereby shared resources, software, and information are provided on demand to computers, smartphones [8], and other similar devices. Cloud computing is also a new style of computing in which dynamically scalable resources are provided as virtualized services [9]. This allows service providers and users to adjust their computing capacity depending on how much is needed at a given time or for a given task.

Mobile computing [10] means using portable devices to run stand-alone applications and/or accessing remote applications via wireless networks.

Increasing the number of mobile applications demands greater resources and improved interactivity for better experience. Resources in cloud platforms such as Amazon EC2 [11], Microsoft Azure [12] and Google AppEngine [13] can remedy the lack of resources in mobile devices. But what does mobile cloud computing really mean? Different people hold different views, and there are several existing definitions of mobile cloud computing.

In [14], mobile cloud computing was first referred to as an infrastructure where data storage and processing could happen outside the mobile device, enabling a new class of applications—especially context-aware mobile social networks. Mobile cloud applications move computing power and data storage away from mobile phones and into the cloud. This brings mobile applications and computing not just to smartphone users but to a broad range of mobile subscribers.

In terms of services [7], mobile cloud computing extends processing and storage beyond devices and seamlessly integrates with other services and sensor data to open up new classes of application.

In [9], mobile cloud computing is defined as an extension of cloud computing in which foundation hardware consists at least partly of

This research is partially supported by Hong Kong RGC under the GRF grant PolyU5106/10E and Nokia Research Lab (Beijing) under the grant H-ZG19. This work is also partially supported by the National S&T Major Project of China under No.2009ZX03006-001 and Guangdong S&T Major Project under No.2009A080207002.

mobile devices. This definition recognizes the opportunity to harness collective sensing, storage, and computational capabilities of multiple networked wireless devices to create a distributed infrastructure that supports a wealth of new applications.

Therefore, our goal in this survey is to eliminate confusion by clarifying the terms and concepts, predicting state-of-the-art mobile cloud computing, discussing future trends, and identifying technical and non-technical obstacles and opportunities.

2 Mobile Cloud Computing

Definitions of mobile cloud computing can be divided into two classes. The first refers to carrying out data storage and processing outside mobile devices [15]. Mobile devices are simply terminals in cloud computing, only intended to provide a more convenient way of accessing services in the cloud. The benefit of this is that storage and computing limitations of mobile devices are avoided, and a new level of security is provided by centralizing maintenance of security-critical software.

The second class of definitions refers to computing where data storage and processing are also carried out on mobile devices [9]. The infrastructure of the cloud is different from that of data-center "pay-as-you-use" cloud computing. Each node is owned by a different user and is likely to be mobile. Using mobile hardware for cloud computing has advantages over using traditional hardware. These advantages include computational access to multimedia and sensor data without the need for large network transfers, more efficient access to data stored on other mobile devices, and distributed ownership and maintenance of hardware.

In this paper, mobile cloud computing is defined as cloud computing extended by mobility and a new ad-hoc infrastructure based on mobile devices. It provides mobile users with data storage and processing services on a cloud computing platform. Using this definition, we clarify two differences between mobile computing and cloud computing.

First, there is an overlap between the two computing paradigms. Cloud computing aims to provide services to users without letting them know where these services are hosted or how they are delivered. Mobile computing aims to support mobility so that users can access resources with wireless technology.

Second, it is possible to build up clouds within mobile devices for data storage and processing. Although mobile devices nowadays are weak in terms of energy, power, storage, and communication, this situation will be improved dramatically with the development of technology. In [9], Hyrax shows that a platform derived from Hadoop [16] can support cloud computing on Android [17] smartphones. It demonstrates the possibility that a cloud composed of mobile devices can provide basic functions and services of cloud computing.

Why do we need mobile cloud computing? There are at least three possible explanations.

First, mobility support is very important in allowing users access to cloud services. Service availability is considered the first obstacle [9] for cloud computing. Mobile cloud computing exploits information about a user's location, context, and requested services and significantly evolves the heterogeneous access management schemes for traditional heterogeneous access scenarios. This is called "intelligent access" [15]. It can noticeably improve user experience.

Second, the capacity of mobile devices has increased dramatically with the growth of mobile users, especially with the prevalence of smartphone users. Each mobile device has storage, computing, sensing, and power resources. By using these resources, applications or services can be easily accessed by users inside and outside the mobile network.

Third, there are still some obstacles to implementing cloud computing, and mobile computing technology can help overcome these obstacles. For example, long WAN latencies are a fundamental obstacle when a mobile device executes a resource-intensive application on a distant high-performance server or cluster. A solution is to instantiate customized service software on a nearby cloudlet [18] and then to use that service over a wireless LAN.

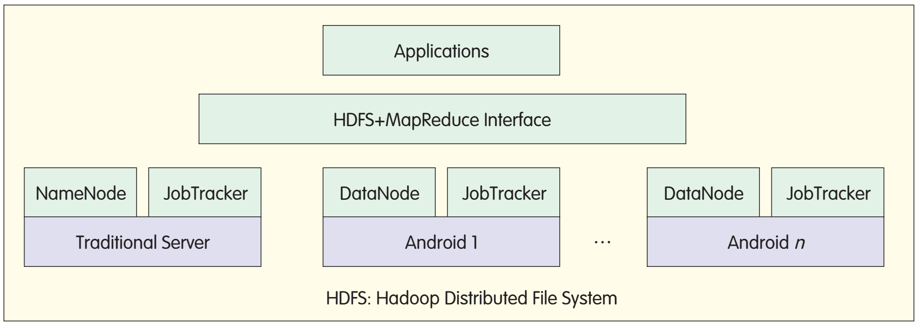
3 Existing Works on Mobile Cloud Computing

According to our definition of mobile cloud computing, there are two key issues: How to build up a mobile cloud using mobile devices and how to access services provided by clouds in a mobile way. Thus, existing work on mobile cloud computing can be classified as that based on the cloud platform, or that based on access schemes. For the former, we investigate existing work on mobile cloud platforms consisting of mobile devices. The Hyrax [9] platform is derived from Hadoop [16] and supports cloud computing on Android [17] smartphones. To the best of our knowledge, it is the only platform comprised of mobile devices. For the latter, we investigate existing access schemes in mobile cloud computing. Mobile cloud computing is still in its infancy, and we investigate Virtual Machine (VM)-based cloudlets [18] and intelligent access schemes [15].

3.1 Hyrax

The goal of Hyrax is to develop a mobile cloud infrastructure that enables smartphone applications with distributed data and computation. Hyrax allows applications to conveniently use data and execute computing jobs on smartphone networks and heterogeneous networks of phones and servers. Research has been focused primarily on implementation and evaluation of mobile cloud computing infrastructure based on MapReduce [19]. MapReduce, Hadoop, and Android should be introduced before describing Hyrax infrastructure.

MapReduce is considered a programming model for processing and



▲Figure 1. Hyrax hardware and software layers.

generating large data sets. A user specifies a map function that processes a key/value pair in order to generate a set of intermediate key/value pairs. It also has a reduce function that merges all intermediate values associated with the same intermediate key. The MapReduce runtime system splits input data, schedules map and reduce tasks, and transfers input and output data to machines running the tasks. There is a master to manage jobs, assign tasks to slave machines, and provide locations of intermediate values to reduce tasks.

Apache Hadoop is an open source implementation of MapReduce that is used by many organizations for large-scale data processing. It is designed to operate data stored in a distributed file system, Hadoop Distributed File System (HDFS). There are four types of processes in Hadoop instances: NameNode, JobTracker, DataNode, and TaskTracker. There is one NameNode and one JobTracker in a Hadoop cluster. NameNode schedules jobs and coordinates sub-tasks among TaskTrackers. A DataNode instance and a TaskTracker instance both run on each worker machine. DataNode stores and provides access to data blocks, and TaskTracker executes tasks assigned to it by JobTracker. Clients access files by first requesting block locations from NameNode and then requesting blocks directly from these locations. The hardware and software layers of Hyrax are described in Fig. 1.

Hyrax ports Hadoop to the Android platform. In Hyrax, each machine runs one instance of NameNode and one instance of JobTracker. DataNode and TaskTracker are run on each phone in separate Android service processes within the same application. Android applications may consist of multiple processes, some of which run as background services. Since DataNode and TaskTracker are run as Android services, they can run in the background of other applications.

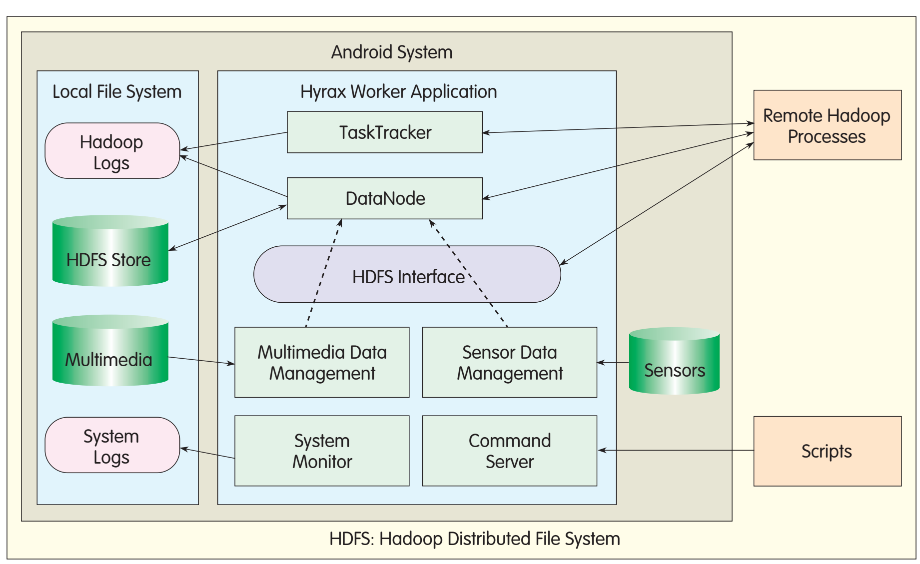
A thread is also spawned to record information about the system load—including power level and CPU, memory, network, and disk I/O statistics—into the local file system. Within the application, a server is run to allow external scripts to control data uploading, kill the program, and check the program status. Fig. 2 illustrates the data interactions between all the software components on each phone.

Hyrax is implemented in a real testbed consisting of a cluster of 10 Android G1 (HTC Dream) phones and 5 HTC Magic phones, each running Android 1.5 Cupcake. Since Android does not support peer-to-peer networking yet, the phones communicate with each other on an isolated 802.11g network via a Linksys WRT54G wireless router with no firmware modifications. NameNode and JobTracker processes are run on a desktop machine connected behind this router via Ethernet. The phones are connected via USB to a controller that executes experiment scripts. These scripts are used to install Hyrax, initialize the cluster, run benchmarks, and collect and post process data.

To determine the advantages and drawbacks of Hyrax, an application was developed on it. The Hyrax multimedia search and sharing application, HyraxTube, allows users to browse videos and images stored on a network of phones and search by time, location, and quality. Quality ratings based on sensor data are generated by periodically executing a MapReduce job. Requests are serviced by reading results generated by the MapReduce job from HDFS. The client interface is implemented as a web application so that it can be used on mobile devices and desktop machines.

3.2 VM-Based Cloudlets

Satyanarayanan et al. [18] present a new vision for mobile cloud computing. They foresee a new world in which mobile computing seamlessly augments users' cognitive abilities via



▲Figure 2. Hyrax worker application component interaction diagram.

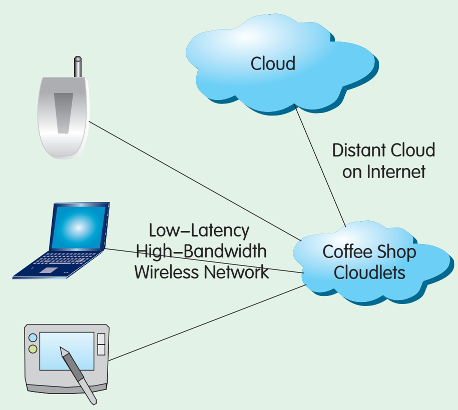
intensive capabilities such as speech recognition, natural language processing, computer vision and graphics, machine learning, augmented reality, planning, and decision making.

The means of achieving this vision is to let a mobile user exploit VM technology to rapidly instantiate customized service software on a nearby cloudlet. That service is then used over a wireless LAN. A cloudlet is a trusted, resource-rich computer or cluster of computers that is well-connected to the Internet and available for use by nearby mobile devices. Mobile devices act as a thin client in a 3-layer cloud computing architecture.

This solution mainly addresses the resource poverty of mobile devices, and long WAN latency is a fundamental obstacle. The predominant view is that WAN latency can hurt usability by degrading the crispness of system response, and it is unlikely to be improved. Bandwidth-induced delays also hurt user experience. The transmission of large data items should occur within a tight user-machine interaction loop.

Real-time interactive response can be achieved through low-latency, one-hop, high-bandwidth wireless access to the cloudlet. Fig. 3 illustrates cloudlets and how they work. A cloudlet resembles a data center in a box; it is self-managing and requires little more than power, Internet connectivity, and access control for setup. Such simplicity of management is similar to an appliance model of computing resources and makes cloudlets simple to deploy in a business premises. Internally, a cloudlet resembles a cluster of multicore computers, with gigabit internal connectivity and high-bandwidth wireless LAN.

Cloudlet infrastructure is deployed in much the same way as Wireless Fidelity (Wi-Fi) access points. The key challenge is to simplify cloudlet management, and a proposed solution is transient customization of cloudlet infrastructure using hardware VM technology. Pre-use customization and post-use cleanup ensure that cloudlet



(a) The cloudlet concept involves proximate computing infrastructure that can be leveraged by mobile devices

	Cloudlet	Cloud
State	Only Soft State	Hard and Soft State
Management	Self-Management	Professional
Environment	Data Center in a Box	Real Data Center
Ownership	Decentralized Ownership by Local Business	Centralized Ownership by
Network	LAN Latency/Bandwidth	Internet Latency/Bandwidth
Sharing	Few User at a Time	Huge Users at a Time

(b) It has some key differences with the basic cloud computing concept

▲Figure. 3. The cloudlet concept.

infrastructure is restored to pristine state after each use, without manual intervention. A VM cleanly encapsulates and separates the transient guest software environment from the cloudlet infrastructure's permanent software environment. Therefore, a VM-based approach is less brittle than alternatives such as process migration or software virtualization. It is also less restrictive and more general than language based virtualization approaches that require applications to be written in a language such as Java or C#.

3.3 Intelligent Access Schemes

Mobile cloud computing relies on permanent connectivity. Intelligent access schemes [15] try to meet this requirement by specific information such as the user's location, context, and request services provided by the so-called mobile cloud controller.

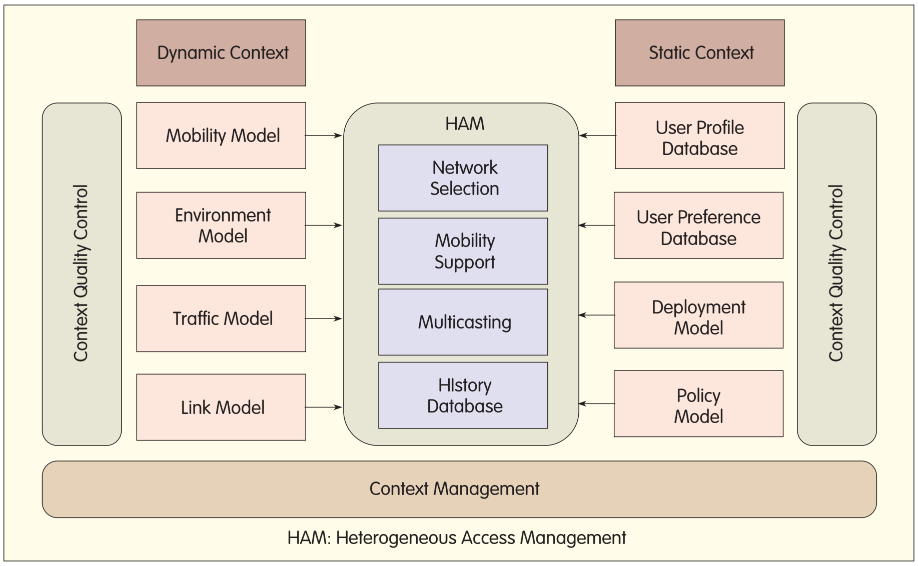
Efficient network access management across different Radio Access Technologies (RATs)—Heterogeneous Access Management (HAM)—is necessary. Radio resources are limited due to physical, technology-specific, and regulatory constraints. Context provided by terminals, network nodes, or sensors deployed in the user's environment are used to significantly reduce wasting scarce radio resources and to efficiently manage wireless access across heterogeneous RATs.

The proposed solution is designed for a heterogeneous access scenario with a wide range of different radio access technologies and context information. The proposed Intelligent Radio Network Access (IRNA) concept takes into account the characteristics and status of each RAT while simultaneously trying to accommodate users who face different environmental conditions in order to provide the best possible end-to-end performance. IRNA is shown in Fig. 4.

The proposed Context Management Architecture (CMA) acquires, processes, manages, and delivers context information. A Context Quality Enabler (CQE) is incorporated into the architecture. The CQE controls the provision of context information according to the requirements of the Mobile Cloud Controller. A Context-Aware Radio Network Simulator (CORAS) is proposed to model context availability, accuracy, and delay, and enable an evaluation to be made on the impact of context relevance, confidence, and quality on simulation results.

4 Conclusions

Wireless and mobile computing technologies provide more possibilities for accessing services conveniently. Mobile devices will be improved in terms of power, CPU, and storage. Mobile cloud computing has emerged as a new paradigm and extension of



▲Figure 4. The concept of Intelligent Access.

cloud computing.

In this survey, terminologies and concepts are clarified, and a definition of mobile cloud computing is provided based on an understanding of underlying technologies and applications. Existing work has also been surveyed and two classes of architecture framework are described for mobile cloud computing.

For future research, there are still some challenging issues related to the models of mobile cloud computing and services. They are:

(1) Mobile devices are constrained in storage and processing capacity. How can efficient use be made of the limited resources for cloud computing?

(2) There are several operating systems for mobile devices, especially mobile smartphones. These include Android, Symbian, iOS [20], Chrome, and MeeGo [21]. Is it possible to provide a general access platform for mobile cloud computing on top of these various OS platforms?

(3) From Hyrax and Cloudlet, mobile devices act not only as clients but also as components of clouds. What is the impact of mobility of mobile devices on performance of the cloud? What are the benefits and disadvantages?

(4) Users in mobile clouds are still exposed to security threats both inside and outside the cloud. What is the best solution for such a mobile environment?

Demand and applications of mobile cloud computing will grow rapidly, and researchers and developers should pay more attention to this area of cloud computing research and development.

References

[1] B. Hayes, "Cloud computing," *Communications of the ACM*, vol. 51, no. 7, pp. 9-11, Jul. 2008.
[2] M. Armbrust et al., "A view of cloud computing," *Communications of the ACM*, vol. 53, no.4, pp. 50-58, Apr. 2010.
[3] R. Buyya et al., "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599-606, Jun. 2009.
[4] "Cloud Computing," Wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/Cloud_computing
[5] Expert Group Report, "The Future of Cloud Computing. Opportunities for European Cloud Computing Beyond 2010," 2010. [Online]. Available: http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf
[6] M. Armbrust, A. Fox, R. Griffith et al. (2009, Feb.) "Above the clouds: A Berkeley view of cloud computing." EECS Dept. University of California, Berkeley. Tech. Rep. UCB/EECS-2009-28. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html
[7] D. Kovachev, D. Renzel, R. Klamma, and Y. Cao, "Mobile community cloud computing: emerges and evolves," *Proc. 1st Intl. Workshop on Mobile Cloud Computing (in conjunction with) 11th Intl. Conf. on Mobile Data Management (MDM'10)*, Kansas, MO, 2010, pp. 393-395.
[8] "Smartphone," Wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/Smartphone/
[9] E. Marinelli, "Hyrax: cloud computing on mobile devices using MapReduce," Master thesis, Carnegie Mellon University, 2009.
[10] Mahadev Satyanarayanan, "Mobile computing: The next decade," *Proc. 11th Intl. Conf. on Mobile Data Management (MDM'10)*, Kansas, MO, 2010.
[11] "Amazon elastic compute cloud (EC2)." AWS. [Online]. Available: http://www.amazon.com/ec2/
[12] Microsoft azure. [Online]. Available: http://www.microsoft.com/azure/
[13] Google app engine. [Online]. Available: http://appengine.google.com/
[14] "The 1st ACM workshop on mobile cloud computing & services: Social networks and beyond," MCS 2010. [Online]. Available: http://www.cs.colorado.edu/~rhan/MCS2010.html
[15] A. Klein, C. Mannweiler, J. Schneider, and H. Schotten, "Access schemes for mobile cloud computing," *Proc. 11th Intl. Conf. on Mobile Data Management (MDM'10)*, Kansas, MO, 2010, pp. 387-392.
[16] Apache Hadoop. [Online]. Available: http://hadoop.apache.org/
[17] Android. [Online]. Available: http://www.android.com/
[18] M. Satyanarayanan, P. Bahl, R. Cáceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," *IEEE Pervasive Computing*, vol. 8, no. 4, pp.14-23, Oct. 2009.
[19] Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *Proc. 6th Symp. on Operating System Design and Implementation (OSDI'0 04)*, San Fransisco, CA, pp.137-150.
[20] Apple IOS. [Online]. Available: http://www.apple.com/iphone/ios4/
[21] MeeGo. [Online]. Available: http://meego.com/

Biographies

Xiaopeng Fan (xp.fan@siat.ac.cn) received B.E. and M.E. degrees in computer science from Xidian University, China. He received his Ph.D. in computer science in 2010 from Hong Kong Polytechnic University. He is currently an assistant researcher at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include cloud computing, wireless communication, and software engineering. His recent research has been focused on data center and mobile data management. He has reviewed for several international journals and conference proceedings.

Jiannong Cao (csjcao@comp.polyu.edu.hk) received his M.Sc. and Ph.D. degrees in computer science from Washington State University, Pullman. He is a professor and acting head of the Department of Computing at Hong Kong Polytechnic University. He is also the director of the department's Internet and Mobile Computing Lab. Before joining Hong Kong Polytechnic University, he taught in the faculty of computer science at James Cook University, the University of Adelaide, and City University, Hong Kong. Dr. Cao's research interests include computer networks, mobile and pervasive computing, parallel and distributed computing, fault tolerance, and middleware. He has co-authored a book on mobile computing, co-edited 9 other books, and published over 300 papers. Dr. Cao is a senior member of the China Computer Federation, a senior member of the IEEE's Computer Society and Communication Society, and a member of the ACM. He is also a member of the IEEE Technical Committee on Distributed Processing, IEEE Technical Committee on Parallel Processing, and IEEE Technical Committee on Fault-Tolerant Computing.

Haixia Mao (lsjessie@polyu.edu.hk) received B.E. and M.E. degrees in remote sensing and geo-informatics at Wuhan University, China. She received her Ph.D. degree in remote sensing and geo-informatics from Hong Kong Polytechnic University in 2010. She is currently a research associate in Hong Kong Polytechnic University. Her research interests include remote sensed image processing and mass geo-spatial data mining in geo-informatics.

Mirroring Smartphones for Good: A Feasibility Study

Bo Zhao[1], *Zhi Xu*[1], *Caixia Chi*[2], *Sencun Zhu*[1], *and Guohong Cao*[1]
(1. Department of Computer Science and Engineering, The Pennsylvania State University;
2. Bell Laboratories, Alcatel-Lucent Technologies)

Abstract: An increasing number of applications and functions are being introduced into smartphones, but smartphones have limited computation ability and battery resources. To enhance smartphone capacity, cloud computing and virtualization techniques can be used to shift the workload from smartphone to computational infrastructure. In this paper, we propose a new framework in which a mirror is kept for each smartphone on a computing infrastructure in the telecom network. With mirrors, the workload can be greatly reduced, and smartphone resources can be virtually expanded. The feasibility of deploying this framework in telecom networks is demonstrated in the protocol design, a synchronization study, and a scalability test. Two applications are introduced to show how computational workload on the smartphone and traffic in the telecom network are significantly reduced using our techniques.

Key words: mirror; cloud computing; smartphone

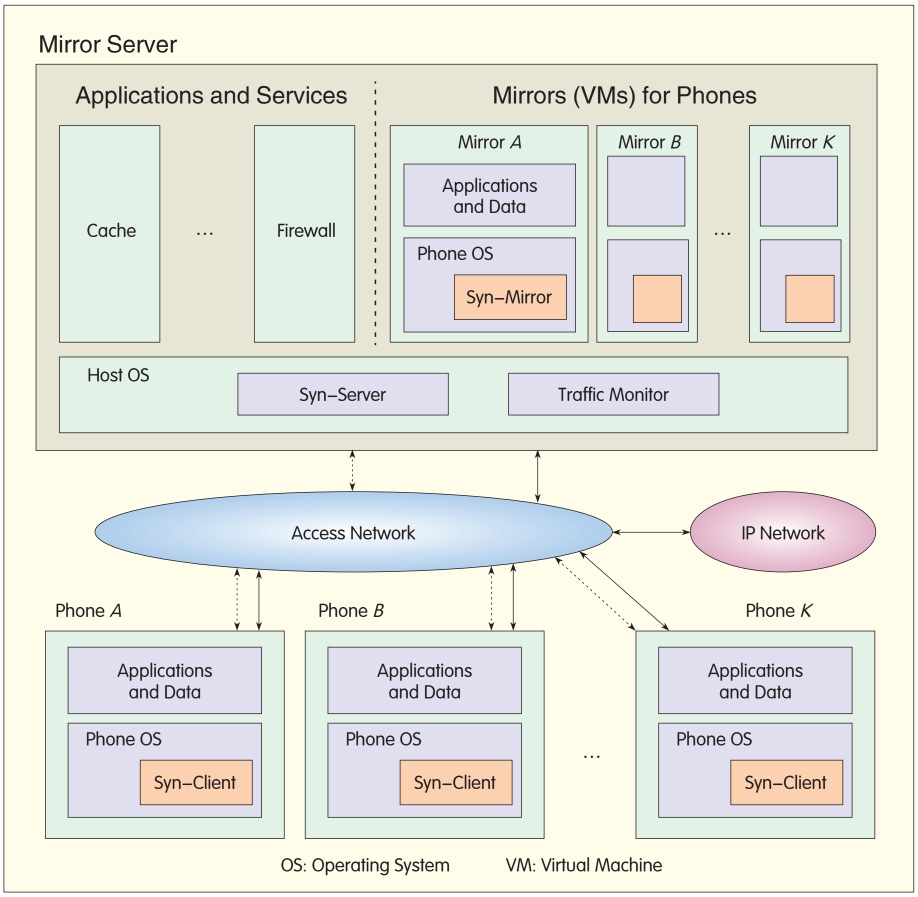
1 Introduction

Smartphones have become more intelligent and powerful. Many complex applications previously used only on PCs are now being developed to run on smartphones. These applications expand the functionality of smartphones and make it more convenient to get connected. However, they also greatly increase the workload on smartphones and result in many more data transmissions between smartphones and telecom networks. Heavy workload and traffic affects both smartphone users and Telecommunications Service Providers (TSPs). For users, heavy workload and traffic drain a smartphone's battery quickly. When we tested an Android Dev Phone 1, scanning a folder with 200 MB files took about 40 minutes and drained the battery by 10%. Uploading a 20 MB file drained the battery by more than 5%.

Recent research [1]-[4] has been done leveraging cloud techniques to help smartphones. An augmented execution model can be used to offload application execution from a smartphone to its clones in the cloud [2]. In an in-cloud antivirus system [3], smartphones can send suspicious files to an antivirus service in the cloud for scanning. This avoids the need for resource intensive scanning to be performed locally on the phone. In addition, Zhang et al. [4] have suggested building elastic applications that can be partitioned into independent function components. These components can execute functions in the cloud instead of on smartphones. Cloudlets [1] are designed for applications that require real-time interactive response, and smartphones are connected to cloudlets via high speed wireless network connections such as Wireless Fidelity (Wi-Fi). In its framework, a smartphone migrates Virtual Machines (VMs) to the cloudlet infrastructure, which provides various services and returns the results to the smartphone. Since cloudlets have much greater computation power compared to smartphones, the execution speed is much faster. However, migrating the whole VM consumes much power and bandwidth.

We propose a new framework in which a mirror is kept for each smartphone on a computing infrastructure in the telecom network. Using mirrors, some of the computation workload can be shifted from a smartphone to its mirror. Since a mirror is synchronized with a smartphone, operations such as file sharing and virus scanning can be performed directly on the mirror. The smartphone's workload is reduced, and its resources are virtually expanded.

This framework design leverages emerging cloud computing and virtualization techniques. On the smartphone side, a client-side synchronization module is deployed to collect user input data and transmit it to the mirror server (a powerful application server). With cloud computing technologies, the mirror server is capable of hosting hundreds of mirrors, and each mirror is implemented as a VM. To keep loose synchronization between the mirror and the



▲Figure 1. The system design.

smartphone, the mirror replays user input on the smartphone in exactly the same order.

To illustrate the feasibility of deploying the proposed framework in 3G networks, we propose a network architecture based on UMTS (a 3G network architecture currently used by T-mobile). Two applications demostrate the benefits of this framework. A data caching application saves power consumption in a smartphone and saves 3G network bandwidth by caching a smartphone's data in a mirror server. Another application is virus scanning in a smartphone.

A prototype of the mirror server was built using Dell PowerEdge 1900. By measuring the protoype's workload when different numbers of mirrors are used, we demonstrate that the scalability of the framework is acceptable. We also evaluate the power consumed by running the synchronization module on an Android Dev Phone 1. The results show that performance overhead on the smartphone is very small.

The remainder of this paper is organized as follows: Section 2 describes the design of our framework. Next we present two applications to show the benefits of our framework in Section 3. Section 4 presents our testbed and experimental results. Section 5 concludes the paper.

2 System Design

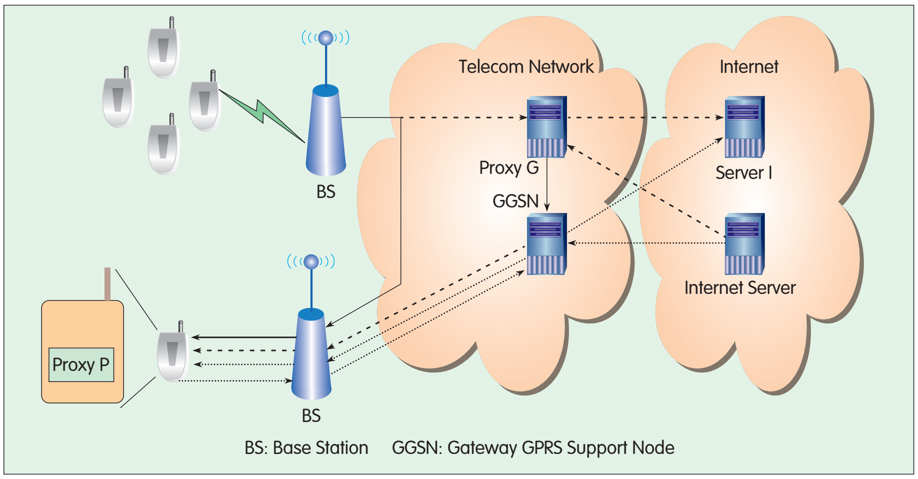
The proposed system provides an architecture for shifting computing from smartphones to their mirrors in the telecom network. Fig. 1 gives an overview of the system. On the smartphone side, a client-side synchronization module called Syn-Client is deployed within the smartphone Operating System (OS). This client collects smartphone input data, including user keyboard input, and transmits it to the mirror server for synchronization. Syn-Client is designed according to specifications of the TSP and manufacturer. On the mirror server side, the mirror server is a powerful application server for maintaining a set of VMs. One VM is a mirror for one smartphone. To keep mirrors and smartphones synchronized, a server-side synchronization module called Syn-Server updates mirrors based on data provided by Syn-Clients and based on network traffic between smartphones and the IP network. Network traffic data is collected by a traffic monitor module. The detailed design and implementation of other modules are not within the scope of this paper.

(1) Location of the Mirror Server

In the proposed architecture, the mirror server monitors all network traffic of smartphones. In this case, traffic from both Internet Service Providers (ISPs) (e.g. Internet websites) and the TSP (e.g. AT&T) is considered. Bluetooth or Wi-Fi connections are not considered in this discussion; we only consider the case where a smartphone is connected through a 3G network.

There are two options for deploying the mirror server: Deploying it in the ISP network (server I in Fig. 2) or in the TSP network (server T in Fig. 3). For the former, the dotted and dashed lines in Fig. 2 represent the message flows with proxy P and proxy G respectively. For the latter, network traffic from TSP and ISP is copied to server T within the telecom network. This requires modifying the existing telecom network, and details of this modification are given in the next section.

In the systems proposed in [4], the cloud that assists smartphones is deployed with the ISP and is connected to smartphones via IP networks. However, in our architecture, if the mirror server is deployed in the ISP network, an additional proxy is needed to collect incoming smartphone traffic and forward it to the mirror server. As shown in Fig. 2, such a proxy can be deployed either beside the Gateway GPRS Support Node (GGSN) in a TSP network or beside the smartphone,



▲Figure 2. Placing the mirror server in the ISP network.

represented as proxy G and proxy P respectively. However, placing a proxy beside the GGSN requires changing the telecom network so that messages can be forwarded and a mirror server can be placed in the TSP network. Taking these issues into consideration, we deploy the mirror server in the TSP network.

(2) Mirror Design

Compared with using PCs, creating mirrors for smartphones has many advantages. First, same model smartphones have the same hardware specifications, default factory settings, OS, and many software applications. Moreover, in the restriction of usage agreement with a TSP, users are not allowed to modify the hardware or OS of their smartphones. Therefore, the mirror server only needs to keep one template of the factory defaults for each model.

To build a new mirror for a smartphone, the mirror server creates a VM with exactly the same hardware/software specifications as the factory default. If the smartphone is in its original state; that is, first-time use, the mirror server can quickly update the mirror by supplying user information. If the smartphone is not in its original state, the mirror server copies software specifications and personal data from the smartphone's storage to the mirror. During the copying process, the state of the smartphone is frozen and no operation by a user is allowed.

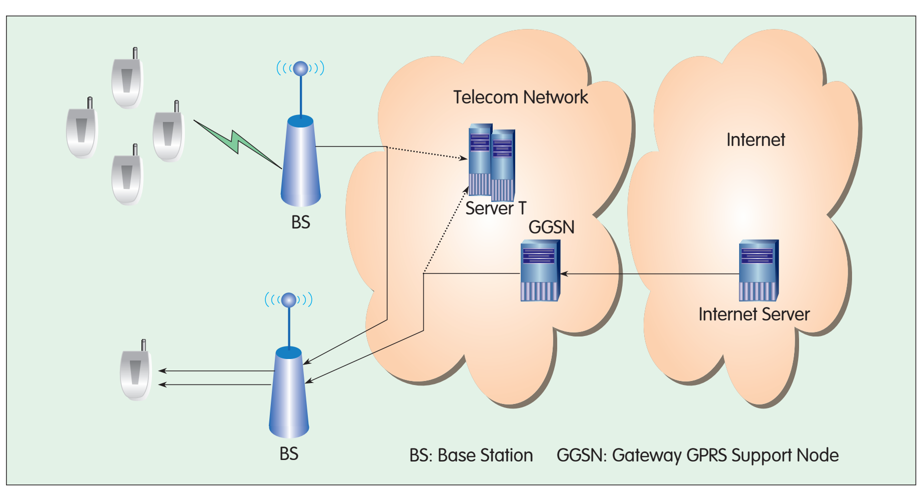
(3) Synchronization Mechanisms

To keep a smartphone and its mirror synchronized through the telecom network, the mirror needs to be updated when the state of the smartphone changes. In this paper, one option is presented for loose synchronization, which requires identical hardware specification, storage (e.g. SD card), OS, and installed application software. In the first stage of our work, we are mainly concerned with maintaining storage consistency between a smartphone and its mirror.

Data replication is an intuitive method of storage synchronization and involves copying every new or updated data item to the mirror. However, this is too expensive when data activity is frequent on the phone. Therefore, a synchronization mechanism that periodically replays smartphone input at the mirror is proposed.

The question for such a design is: Why is the data storage uniquely determined by user input? We noticed that changes in a smartphone's state are always triggered by user inputs (e.g. keyboard typing) or network inputs (e.g. arrival of data packets) and that most applications are deterministic. Given a list of inputs, the application generates deterministic outputs. In this case, if the same list of inputs is applied to a smartphone and its synchronized mirror, both generate the same output and are synchronized again after applying the inputs. Note that in our design, to save communication overhead, only user inputs are forwarded to the mirror. The mirror caches the network inputs to a smartphone in the previous synchronization period so that later it can directly feed the cached data to the mirror. Also, the output from the mirror execution does not actually leave the mirror server. The global clock of the telecom network is used to timestamp all inputs/outputs. For example, on the smartphone side, the user pushes a button to open an email client application, types an email, and sends it by clicking an icon on the screen. On the mirror side, if the "pushing," "typing," and "clicking" actions are exactly replayed in order, the mirror sends the same email and both have the same archived email in storage.

It can therefore be assumed that, by



▲Figure 3. Placing the mirror server with the TSP.

D:\EMAG\2010-05-26/VOL8\COVER.VFT——5PPS/P15

applying exactly the same inputs to a smartphone and its mirror in the same order, the storage of a phone and its mirror will be synchronized. This assumption may be untrue for certain applications, such as those that generate a random number. However, it is suitable for most email, IM, web browser, VoIP, and MMS/SMS smartphone applications as well as this mirror service. Therefore, all inputs to the smartphone are monitored and later applied to the mirror in order. Periodic synchronization improves transfer efficiency and reduces power consumption greatly by sending the inputs in a batch.

(4) Network Design

The mirror server resides in a 3G network as an application server and supports services such as security, storage, and computation delegation. It is easy to configure the nodes of a 3G network to route all incoming/outgoing smartphone messages (such as session control messages, data packets, and other 3G signaling protocols) to traverse the mirror server without any interaction with the smartphone. It is not necessary to modify any protocol or software of the network nodes.

Take UMTS for example. Each subscriber to the mirror service has their user profile updated in the Home Subscriber Server (HSS) and their user policy and mirror service profile registered at the Call Session Control Function (CSCF).

The HSS periodically updates the Serving GPRS Support Node (SGSN) with changes of user information; thus, SGSN becomes aware of the mirror service for a user. When SGSN receives a message from/to a user, it either forwards the message to CSCF or forwards it to the mirror server depending on whether the message belongs to a service with session control.

Technical report [5] details the deployment of a mirror server in a UMTS network. In [5], two sample message flows of typical telecom services are presented to show that the mirror server is compatible with a UMTS network, and only minor changes are needed.

3 Applications

The proposed framework can support many applications. In this section, two are identified.

3.1 Data Caching Application

Increasingly, people are using their smartphones to share multimedia data with friends. Although data downloading/uploading speed between smartphones and BS has significantly increased in recent years, receiving and sending bulk data still takes time and consumes a lot of power.

File sharing on smartphones currently has many weaknesses. For example, if the receiver runs out of battery or is downloading very slowly in the current location (weak signal), the download may have to be abandoned. Also, if a user sends the same file to different people in a short period of time, for each send the smartphone must do a separate file upload, wasting a lot of bandwidth and power.

Based on the proposed framework, we present a data caching application. It is deployed on the mirror server, which brings more flexibility to downloading and significantly saves battery and upload time.

(1) Bulk Data Downloading

For bulk data downloading, data caching provides a temporary storage service. When a user wants to download a file from the Internet, the caching service first downloads the file to its temporary storage. Once the downloading is complete, the caching service sends a "file-cached" notification to the user via synchronization messages. This is kept for a period of time depending on the TSP policy. When the user decides to download the cached file, they obtain it directly from the mirror server. Moreover, according to the synchronization mechanism, once the smartphone starts to download a file from the mirror server, its corresponding mirror also downloads the file from the mirror server. Therefore, no extra storage synchronization is necessary.

(2) Bulk Data Uploading

For bulk file uploading, data caching uses the smartphone's mirror on the server. The mirror has the same storage as its corresponding smartphone and can replay all actions performed on the smartphone. Thus, if a user wants to send a bulk file to multiple people, there is no need to upload the file multiple times through the wireless link. Instead, data caching can send the file directly from the mirror to multiple people. Compared with uploading from smartphone, synchronization consumes much less battery power, especially when the file size is large.

3.2 Antivirus Scanning Service

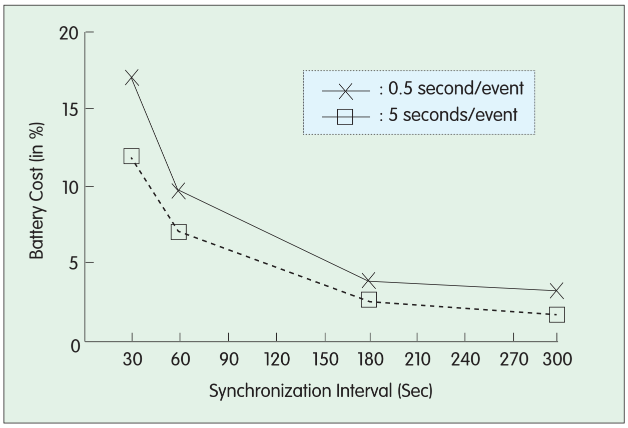
Recent research [6], [7] shows there is an influx of worms, viruses, and other mobile phone malware on 3G networks. In a virus scan, the scanner reads files, compares them with virus signatures, and returns a result to the user. The scan process is CPU and I/O intensive, and in a smartphone, is inconvenient and drains the battery quickly.

In the proposed framework, an antivirus scanner can be deployed as a service on the mirror server, and the scanner can access the file system on the mirrors. Furthermore, a hook application is installed on the phone. When a user launches the hook on the smartphone, it sends a request with parameters to the mirror server. Instead of scanning the real files on the smartphone, the scanner scans the synchronized mirror. The result is sent back to the smartphone. Because the smartphone and its mirror are synchronized, scanning the mirror generates the same result as scanning the smartphone.

The benefits of shifting antivirus scans are threefold as follows: Battery power of the smartphone is saved; CPU and I/O intensive workloads are offloaded so that performance of the smartphone is not affected; and the scan speed on the mirror is much faster than that on the smartphone with limited capabilities.

4 Evaluation

(1) The Prototype



▲Figure 4.
The battery cost of periodic sending over 6 hours with different intervals.

Because UMTS 3G networks are very complicated with multiple protocols and many different nodes, only big telecom equipment vendors such as Nokia, Alcatel-Lucent, and Ericsson have real prototypes. In order to evaluate the performance, scalability, and feasibility of our framework, we set up a simplified prototype. Android Dev Phone 1 was used as the smartphone, and the Dell PowerEdge 1900 was used as the mirror server. The PowerEdge 1900 has two Quad Core Intel Xeon 1.60 GHz CPUs, 8 GB Memory, 320 GB Disk, and Ubuntu 8.04 Desktop OS. The mirror server was connected to the IP network and could be accessed by the smartphone through T-mobile's 3G network. Because we could not route or collect incoming and outgoing smartphone messages in T-mobile's 3G networks, for the synchronization feature in this prototype we only replayed the user input data on the mirror so that the mirror took the same tasks as the smartphone.

A syn-client module was developed for the smartphone to collect input data in real time and periodically transfer it through the 3G network to our mirror server. On the mirror side, Android emulator was the mirror.

(2) Synchronization and Battery Cost

In the proposed framework, the Syn-Client module inside the smartphone sends the input information to the mirror periodically. The battery cost of periodic sending is determined by the input frequency and the time interval between two sending operations. To measure the power consumption of different input frequencies and different sending intervals, the power consumption of periodic sending was measured in two cases. In the first case, the smartphone was very active and generated two events every second. In the other case, the smartphone only generated one event every five seconds. In both cases, the total battery cost was measured with different sending time intervals over a 6 hour period.

Battery cost is shown in Fig. 4. With the same frequency of event generation, shorter sending intervals consume more battery power. At the same sending frequency, more input events cause more data to be sent and more battery power to be consumed.

To determine the benefit of downloading/uploading, the battery cost of synchronization was compared with file downloading/uploading using ftp and a 20 MB file. For periodic synchronization, a case was chosen in which the user generates 0.5 events per second on average and the smartphone synchronization period is 60 seconds. Table 1 shows that the average power consumption of a smartphone in standby and that using synchronization are very close. Compared to these, file downloading/uploading using ftp consumes five times more power.

To determine the benefit of our solution in antivirus scanning, SMobile VirusGuard [8] was installed on an Android Dev Phone 1 and used to scan folders of different sizes. All folders were filled with Android Package (APK) files used for application installation. As shown in Table 2, the time and battery needed for a scan increases as the folder size increases. When the folder size is 200 MB, it takes more than 39 minutes and 10% battery to complete the scan. During the scan, a user can explicitly feel the smartphone slow down. Table 2 also shows that much less time is spent scanning the same folders on a PC using Symantec Endpoint Protection software.

(3) Scalability

To evaluate the scalability of the proposed framework, we studied how many mirrors the server could support with different workloads. In this experiment, a web browsing service was used as an example to evaluate scalability.

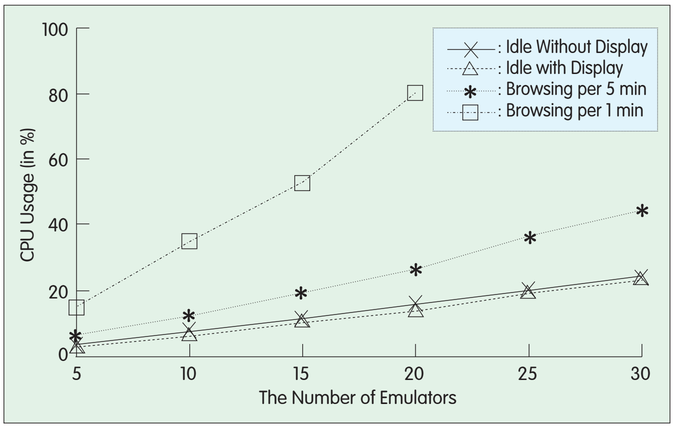
Fig. 5 illustrates CPU usage versus different numbers of emulators together with different workloads. In general, the mirror server can support at least 20 mirrors with heavy workload (browsing the Web every 1 minute) or more than 30 mirrors with light workload (opening a Web page every 5 minutes). In this case, CPU usage increases almost

▼Table 1. Average power consumption per minute

	Standby	Upload	Download	Synchronization
Power (%)	0.09	0.733	0.588	0.117

▼Table 2. The time taken and battery cost of performing antivirus scanning

Folder Size (MB)	SMobile on Smartphone		Symantec on PC
	Time (sec)	Battery (%)	Time (sec)
10	120	2	12
50	600	3	31
100	1140	6	57
200	2340	10	110



◀Figure 5. The scalability for running multiple emulators on a server.

linearly with the number of mirrors.

5 Conclusion and Future Work

In this paper, we propose a new framework that uses cloud computing and VM techniques to shift the workload from smartphones to a computational infrastructure in telecom networks. This greatly reduces the workload of a smartphone and virtually expands a smartphone's resources. The feasibility of deploying such a framework is shown in the protocol design, a synchronization study, and a scalability test. In future research work, we will study different synchronization mechanisms and determine which mobile applications can benefit from the proposed framework. We will build the whole system and evaluate its performance in a real 3G network. We also intend to investigate more complex scenarios with additional incoming traffic through Bluetooth and Wi-Fi connections.

References

[1] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," *IEEE Pervasive Computing*, vol. 8, no. 4, pp. 14-23, 2009.

[2] B. G. Chun and P. Maniatis, "Augmented smartphone applications through clone cloud execution," in *Proc. HotOS XII*, Monte Verità, Switzerland, 2009.

[3] J. Oberheide, K. Veeraraghavan, E. Cooke, J. Flinn, and F. Jahanian, "Virtualized in-cloud security services for mobile devices," in *Proc. MobiVirt 2008*, Breckenridge, CO, pp. 31-35.

[4] X. Zhang, J. Schiffman, S. Gibbs, A. Kunjithapatham, and S. Jeong, "Securing elastic applications on mobile devices for cloud computing," in *Proc. CCSW 2009*, Chicago, IL, pp. 127-134.

[5] B. Zhao, Z. Xu, C. Chi, S. Zhu, and G. Cao. "Mirroring smartphones for good: A feasibility study," CSE Technical report, presented in Mobiquitous'2010. [Online]. Available: www.cse.psu.edu/~bzhao/mirror report.pdf

[6] Z. Zhu, G. Cao, S. Zhu, S. Ranjan, and A. Nucci, "A social network based patching scheme for worm containment in cellular networks," in *Proc. INFOCOM 2009*, Rio de Janeiro, Brazil, pp. 1476 - 1484.

[7] B. Zhao, C. Chi, W. Gao, S. Zhu, and G. Cao, "A chain reaction dos attack on 3G networks: Analysis and defenses," in *Proc. INFOCOM 2009*, Rio de Janeiro, Brazil, pp. 2455-2463.

[8] "Smobile virusguard for google android." [Online]. Available: www.smobilesystems.com

Biographies

Bo Zhao (bzhao@cse.psu.edu) is a Ph.D. Student in the Department of Computer Science and Engineering, The Pennsylvania State University. He received his B.S. degree in computer science from the Peking University in 2003. His research interests include 3G network security, power saving of the telecom networks, and 3G network services.

Zhi Xu (zux103@cse.psu.edu) is a Ph.D. student in the Department of Computer Science and Engineering, The Pennsylvania State University. He received an M.S. degree in Computer Science from the University of Ottawa in 2007 and a B.Eng. degree from Beijing University of Aeronautics and Astronautics in 2004. His research interests include mobile network security and online social network security.

Caixia Chi (chic@alcatel-lucent.com) is a researcher at Bell Labs. She received her Ph.D. degree from the Institute of Applied Mathematics, Chinese Academy of Sciences in 2000. Her research focuses on optical network, network modeling and optimization, algorithm design, and graph theory.

Sencun Zhu (szhu@cse.psu.edu) is an associate professor in Department of Computer Science and Engineering, The Pennsylvania State University. He received his Ph.D. degree in Information Technology from George Mason University in 2004, an M.S. degree from University of Science and Technology of China in 1999, and a B.S. degree from Tsinghua University, China in 1996. His research interests include network and systems security, software security, and social network security.

Guohong Cao (gcao@cse.psu.edu) received the M.S. and Ph.D. degrees in computer science from Ohio State University. Since then, he has been with the Department of Computer Science and Engineering at The Pennsylvania State University, where he is currently a professor. His research interests include wireless networks and mobile computing. He has published over one hundred papers in the areas of sensor networks, wireless network security, data dissemination, resource management, and distributed fault-tolerant computing. He has served on the editorial board of the IEEE Transactions on Mobile Computing and IEEE Transactions on Wireless Communications and has served on the program committee of many conferences. He was a recipient of the NSF CAREER award in 2001. He is a fellow of the IEEE.

Roundup

Tiscali Chose ZTE to Develop Its IMS Network

ZTE Corporation has announced the signing of an agreement with Tiscali for the development of an IMS network. ZTE and Tiscali plan the launch of the first three commercial sites in Cagliari, Roma and Milano by the end of May and the migration of Tiscali subscribers to the new IMS platform by the end of the year.

ZTE will install a single platform in place of the two platforms currently utilised by Tiscali. By implementing the ZTE solution, Tiscali will make its network management easier and more efficient, reducing energy consumption and optimising operational costs.

"We are glad to build our partnership with ZTE Corporation," said Luca Scano, Tiscali General Director, "With the introduction of this unique IMS platform, we will manage to optimise time, resources and results on all our IP communications services, giving our customers a big advantage in terms of price and quality of service." (ZTE Corporation)

A Cloud–Based Virtualized Execution Environment for Mobile Applications

Shih–Hao Hung, Tei–Wei Kuo, Chi–Sheng Shih, Jeng–Peng Shieh, Chen–Pang Lee, Che–Wei Chang, and Jie–Wen Wei
(Department of Computer Science and Information Engineering, National Taiwan University)

Abstract: Smartphones and cloud computing technologies have enabled the development of sophisticated mobile applications. Still, many of these applications do not perform well due to limited computation, data storage, network bandwidth, and battery capacity in a mobile phone. While applications can be redesigned with client–server models to benefit from cloud services, users are no longer in full control of the application. This is also a serious concern. We propose an innovative framework for executing mobile applications in a virtualized cloud environment. With encryption and isolation, this environment is controlled by the user and protected against eavesdropping from cloud providers. We have developed efficient schemes for migrating applications and synchronizing data between execution environments. Performance and power issues within a virtualized execution environment are also addressed using power saving and scheduling techniques that enable automatic, seamless application migration.

Keywords: smartphone; cloud computing; mobile network; virtualization; collaborative computing; energy–saving; scheduling

1 Introduction

Smartphones and cloud computing technologies have enabled the development of sophisticated and pervasive applications. Yet applications on the latest generation of smartphones are limited by battery power, computation speed, and memory size of the smartphone as well as bandwidth in the wireless network [1]. Offloading some of the application workload from the smartphone to a server can save execution time and conserve energy [2]. Today, many cloud–based services are available that have low amortized operation costs [3]. This client–server model has worked quite successfully over the years. Remote execution of mobile applications can save much power [1], but program partitioning is required. Furthermore, as cloud–based services become popular, security and privacy issues are also raised. Facebook, for example, has long been criticized for the privacy risks associated with its site. The following issues in cloud services are of concern to users and application developers:

- Application redesign and deployment: Existing applications have to be partitioned, and the server part has to be deployed. This requires extra development and management effort and cost.
- Network condition and service availability: The quality of a cloud service depends largely on network condition. This raises issues for applications that still need to perform well in poor network connections.
- Control of applications: If a service provider performs computation, users are no longer in full control of their applications. Developers and users can be trapped by proprietary interfaces and could be treated unfairly by a provider.
- Privacy of personal data: User privacy can easily be violated by a service provider. It is not uncommon for service providers to access user data, and it is nearly impossible for users to monitor how this data is used.
- Information security: Offloading workload increases security risks because data is propagated over the Internet and stored with the service provider. During transmission, data is open to eavesdropping by attackers or the service provider.

To address these issues, we propose a framework in which a user can create a virtualized execution environment (virtual environment) in the cloud for running mobile applications. Unlike a client–server model, redesigning

This work was supported in part by a grant from the National Science Council under No. 98–2220–E–002–020, 99–2220–E–002–026, and 95–2221–E–002–098–MY3.

applications is not necessary—a user can have an existing application running on a physical device or in a virtual environment. This approach allows a user to control the deployment and execution of applications, and all that is needed is a trustworthy service provider to host the virtual environment. This is far more practical than verifying a growing number of service providers.

For mobile applications, the communication cost of migrating a process [4] or a virtual machine [5] can be prohibitively high. To accelerate live migration for mobile applications, Android [6] was used as our case study, and several strategies were developed. First, an innovative coarse-grain application migration mechanism was developed based on application-level state-saving mechanisms available in the Android operating environment. Second, types of application data were further categorized to determine necessity and priority for data synchronization. Section 2 discusses the framework and the virtual environment we propose for enhancing Android applications.

So that mobile devices and cloud servers can work together seamlessly on one application, the operating environments on the devices need a collaborative mechanism. In our approach towards collaborative computing [7], the operating environment is enhanced to play an active role in scheduling and distributing workloads. To better utilize the heterogeneous resources on mobile devices and on cloud servers, middleware is placed between mobile operating systems and hardware to support run-time workload migration. To ensure portability, supported Application Programming Interfaces (APIs) are extended beyond what is defined by OpenCL [8] for run-time workload migration. System architectures and design of major components are covered in Section 3.

As we studied the performance and power efficiency of mobile applications in a virtualized environment, we found that energy consumption in virtualized environments can be greatly reduced when power-saving techniques are applied [9]-[11]. Our previous work provides an energy efficient mapping technique for virtual cores [12] and is summarized in Section 4.

2 A Virtual Environment for Android Applications

In this section, we introduce the framework that allows Android users to offload applications to virtual environments in the cloud. The framework automates the creation of a virtual environment and migrates live Android applications faster than traditional methods. Compared to a conventional scheme, this approach does not require developers to redesign their applications, and several effective techniques are proposed for migrating applications and data over a mobile network. Security and privacy measures are included.

2.1 Related Works

We researched the process of partitioning applications—which is performed to take advantage of remote execution. Spectra [13] can be used to monitor resource availability and dynamically determine the best remote execution plan for an application. Cyber foraging [14] surrogates can be used to improve performance of interactive applications and distributed file systems on mobile clients. Mobile Assistance Using Infrastructure (MAUI) [15] reduces the amount of programming needed by using a combination of code portability, serialization, reflection, and type safety to automate program partitioning. If the application is not partitioned, process migration and virtual machine migration are two common methods for migrating the execution of a live application across the network. An Internet Suspend/Resume (ISR) system [16] emulates the capabilities of suspend/resume functions in a computer system and migrates the system by storing a snapshot image of a virtual machine in a distributed storage system.

2.2 The Android Smartphone Framework

This paper introduces a framework for virtualizing Android applications [6]. According to AndroLib [17], by July 2010, there were more than 100,000 Android applications in existance. To optimize smartphone applications, the Android system may kill a process if the system is short of memory. Thus, during its activity lifecycle, an application is designed to save application state data whenever it is switched to the background. When an application receives a request to suspend, its onPause() method is called so that unsaved changes in its state are committed to persistent data and animations or other operations are stopped that may be consuming CPU. When a user gives attention to an activity or a new intent is given, its onResume() method is called. Since the application may be killed by the Android operating environment during suspension, the onResume() method should include instructions to restore the application state before the activity gets ready to receive input from the user. Developers are advised to use the pause-resume scheme provided by Android to save application states in the persistent storage so that the application can resume later. Since most Android applications follow this programming paradigm, we leveraged pause-resume to design our own application migration scheme.

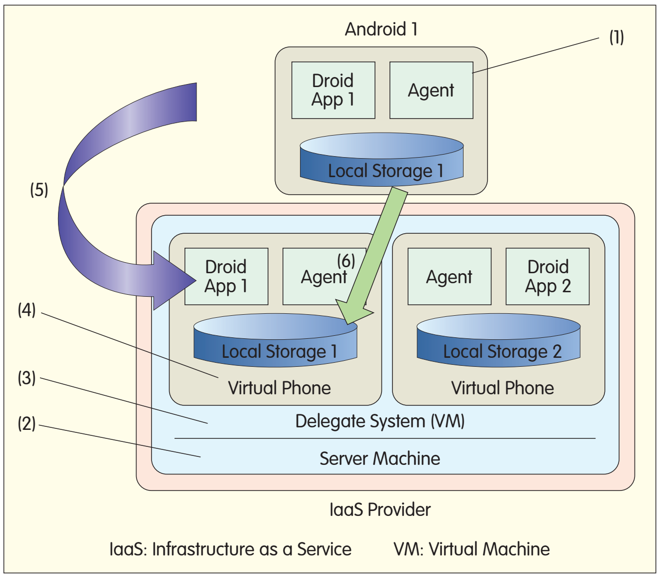
2.3 Our Proposed Framework

In our migration framework (Fig. 1) the following procedures are necessary for creating a virtual environment:

(1) An agent program is installed: The user installs and runs an agent program that automates the rest of the procedures. The agent also provides an interface and applications for interacting with the virtual environment.

(2) A delegate system is allocated: The agent allocates a delegate system to host the virtual environment by subscribing to a virtual machine of an Infrastructure as a Service (IaaS) provider. The delegate system may host multiple virtual environments to save operation cost.

(3) A virtual environment is established: The agent establishes a virtual environment (a virtual phone) on



◀Figure 1. Creating a virtual environment.

the delegate system to emulate an Android phone. For compatibility, the virtual phone needs to emulate all the details of a physical Android device as much as possible.

(4) The operating environment is cloned: The agent uses a standard image stored in the delegate system to create a fresh virtual environment. It then copies the applications and data from the physical phone. An exact clone of the operating environment should ensure compatibility for applications that require vendor-specific libraries or system services.

(5) Applications are migrated: The agent on the physical phone receives commands from the user and communicates with the agent in the virtual environment to control operation of the virtual environment. The user (or the application itself) may request the agent to migrate the application between the two phones.

(6) Applications and user data are synchronized: The agent programs on both phones collaborate to keep the application packages and user data consistent and coherent on both phones. Since continuous mirroring of files would generate a large amount of network traffic, synchronization policies and protocols are critical.

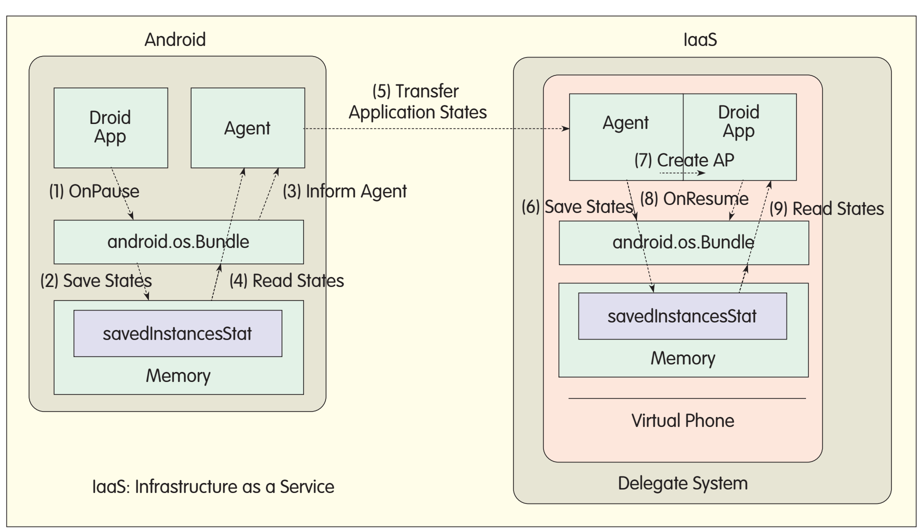
2.4 Migrating an Application

Our application migration mechanism leverages the Android framework to minimize the amount of data needed to migrate a live application. An application is paused on one device, state data files (which are saved by the application as it enters the pause state) are sent, and the application is resumed on another device. Because the state data files are usually small, there is low migration overhead.

The procedure for migrating an application is illustrated in Fig. 2. On the left hand side: (1) The agent sends a signal to the application instructing it to use the onPause function. (2) The application saves its states using the onPause function and (3) informs the agent when the states are saved. (4) The agent reads the states and (5) sends the states to the agent on the other side. Then, on the right-hand side: (6) The agent saves the states and (7) starts the application (or copies the application from the other side if it does not exist). (8) The application resumes by calling the onResume function and (9) resumes the execution after restoring the application state.

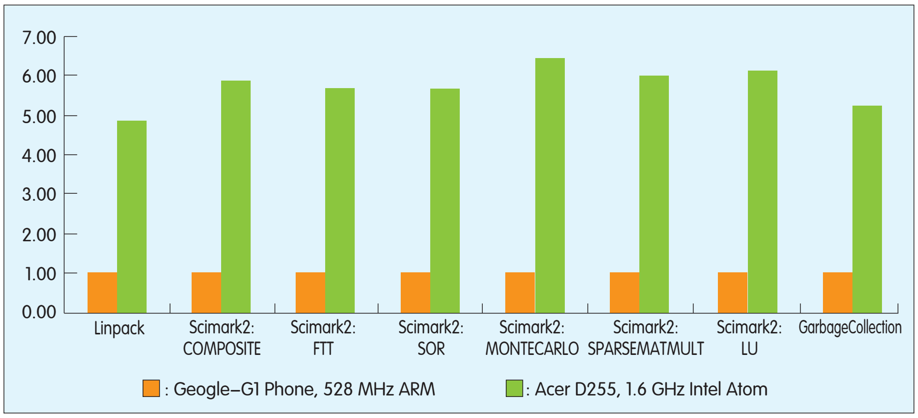
2.5 Security and Privacy Measures

Since a virtual environment may operate in a public cloud, it is important to protect user data with a secure end-to-end communication channel. Choosing a trustworthy IaaS provider is also critical for deploying in a virtual environment. A strong authentication and encryption scheme is needed for secure storage of private keys and for creating a secure channel for exchanging master encryption keys. Communications can occur via a secure Virtual Private Network (VPN) channel. For a stronger level of trust, a Trusted Platform Module (TPM) could be incorporated to enhance security on the server system. A TPM provides hardware mechanisms for storing encryption keys and performs cryptographic operations on sensitive data. To further prevent intrusion from a service provider or attackers, sensitive data in the memory and in storage could be encrypted. Since files are encrypted and hashed, attackers from another virtual machine on the same



▲Figure 2. Migrating an Android application.

D:\EMAG\2010-05-26\VOL8\COVER.VFT——5PPS/P21



▲Figure 3. Performance comparison between a physical phone and a virtual phone.

host or in the middle of the network will find it harder to retrieve and manipulate file contents. TPM could be used to store the master encryption keys and perform encryption procedures to keep the keys safe from intruders.

2.6 Evaluation

In our experiment, Intel Atom–based system host Android–x86 was used as a virtual environment in the cloud. As shown in Fig. 3, the 1.6 GHz Intel Atom processor was already 4.9 to 6.4 times faster than the 528 MHz ARM processor in the Android phone. The results suggest that virtual environments powered by servers with low–end processors still provide sufficient performance for average applications. Migrating an application using a traditional approach [4] would require transferring the state of the entire environment by taking a snapshot of the memory. Assuming an Android system has 512 MB of memory, it would take more than one hour to transfer the snapshot. With our approach, it would only take milliseconds to transfer the state files saved by Android applications.

3 Automatic Task Scheduling and Offloading in a Virtual Environment

A mechanism on the application level can be used to migrate workload. This mechanism is a better fit for a run–time environment that supports application level suspension and resuming. At the same time, the workload can also be migrated on a system level. The system has accurate information on available resources such as network connectivity and computation capacity. On the system level, middleware (called hypervisor) located between the operating system and physical hardware is responsible for allocating resources to processes, dispatching processes to allocated resources, and synchronizing computation results between mobile devices and cloud servers.

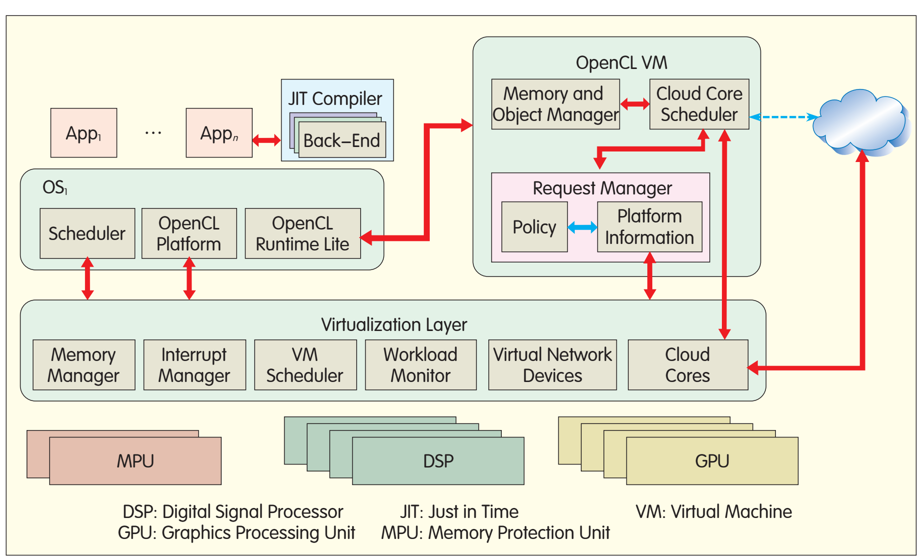
3.1 System Architecture

To support system–level workload migration, OpenCL support can be provided for system software. Fig. 4 illustrates the system architecture for hypervisor with cloud core support. In the architecture, virtual processors (cloud cores) are introduced for the purpose of representing the computation resource on cloud servers.

There are three major source management components in virtualization layers: Virtual Machine (VM) scheduler, memory manger, and interrupt manger. VM scheduler schedules virtual machines on the platform. There are at least two virtual machines: Application VM (for supporting application services and user interactions) and OpenCL VM (for managing OpenCL workload including the workloads on cloud cores). When it is suitable to execute workload in the cloud core, the corresponding workload will be migrated to the cloud core. Otherwise, the workload is executed on local processors. VM scheduler takes into account available resources worldwide and schedules virtual machines according to QoS requirements. The interrupt manager dispatches interrupts and accepts interrupt requests. Interrupts that are managed include those for local cores and cloud cores as well as for peripherals. The memory manager allocates memory to virtual machines.

3.2 Virtual Network Devices

Reliable communication between mobile devices and server is important. Because of the dynamics of wireless networks and communications networks, we designed a



▲Figure 4. System architecture for mobile virtualization with cloud core support.

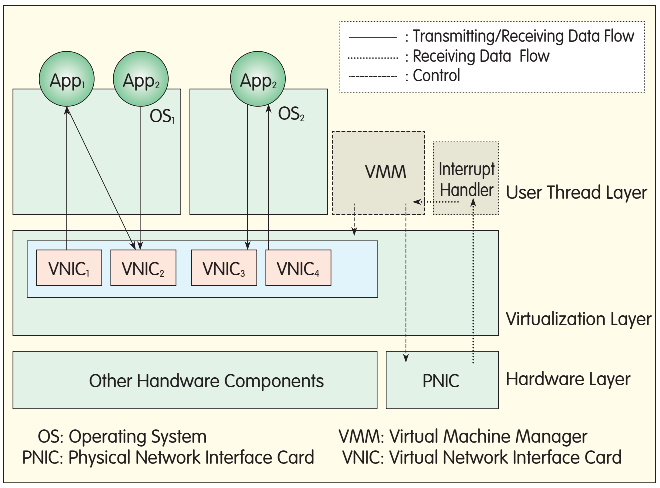


Figure 5. Architecture for virtual network devices.

probabilistic-guaranteed connection via virtual devices. In our framework, resource management is implemented with virtualization technology because different QoS types can be integrated in the virtualization layer. Different applications have different QoS requirements; for example, audio quality and delay are important to VoIP, while frame rate is important to streaming video. Meeting different QoS requirements is not easy. But different QoS requirements can be integrated if a QoS controlling mechanism is implemented on the virtualization layer. A QoS requirement is a special bandwidth requirement. So all QoS requirements can be managed using a bandwidth management concept. For this reason, the QoS framework is implemented with virtualization technology.

Fig. 5 illustrates the operation of virtual network devices. The user thread layer provides services to users or manages hardware resources in the system. Threads are programs such as Operating System (OS) and stand-alone programs that provide services to users. The user thread layer includes all user threads. For transmitting, a network bandwidth allocation mechanism is provided according to service bandwidth requirements. In a dynamic network environment, our QoS framework for transmitting provides bandwidth guarantee with probability. For receiving, a resource protection mechanism is provided to set a maximum data receiving size for services. To verify this framework, we conducted some experiments. The results show that for transmitting, the framework guarantees resource allocation with probability, and for receiving, the framework guarantees resource protection. An extra run-time monitoring mechanism was also added for better performance.

Fig. 6 shows evaluation results of three virtual network devices. Virtual Network Interface Card 1 ($VNIC_1$) has highest priority, requesting 600 kbit/s; $VNIC_2$ has second highest priority, requesting 100 kbit/s; and $VNIC_3$ has the lowest priority, requesting best effort. $VNIC_1$ has the highest probability of transmitting at its requested QoS, and $VNIC_2$ can only transmit at its QoS level when $VNIC_1$ transmits at its QoS level.

Virtual network devices lay the foundation for reliable communication channels in the proposed framework. When a message needs to be transmitted with high robustness, a virtual network device for low bandwidth and high probability is used.

3.3 Cloud Core

The cloud core provides a workload migration mechanism between mobile devices and cloud servers. In particular, it migrates workload implemented with OpenCL API. In OpenCL, basic work units are defined by a user-defined function call. In other words, a user can choose to allocate one function call to be executed on a particular processing element. With this definition, the cloud core can migrate one user-defined function to be executed on the cloud server. In this framework, a client-server model is used to migrate workload to cloud servers.

4 Power Saving for Mobile Applications in a Virtual Environment

Here, we discuss computing resource mapping and the energy relationship between virtual cores and physical cores when there are timing constraints in executing tasks. As opposed to the preceding work, we are interested in Digital Video Server (DVS)

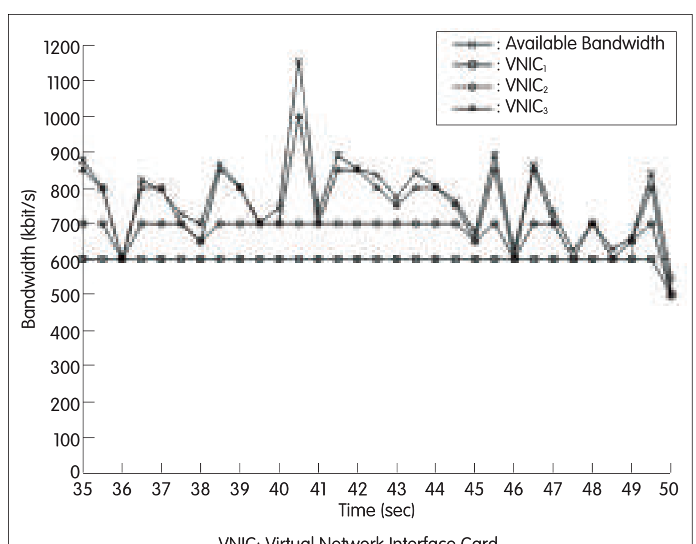
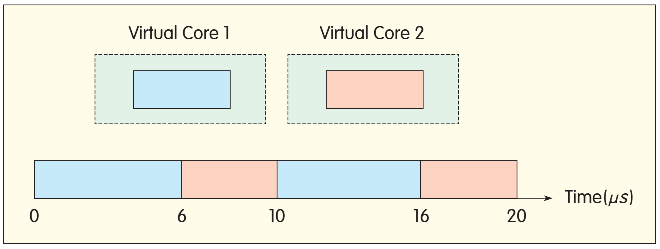
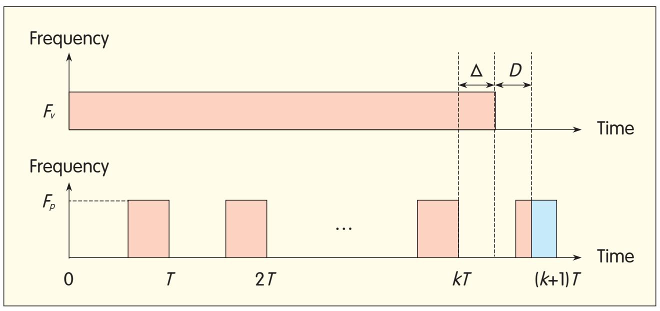


Figure 6. Probabilistic QoS guarantee by virtual network devices.



◀Figure 7. Two virtual cores over one physical core.



◀Figure 8. The delay of the response time incurred while the virtual core is emulated on a physical core.

implementation issues in virtual cores. Adjusting the operating frequency of a virtual core might result in proper dynamic voltage scaling (or even turning off) of some physical cores and/or might trigger the service adjustment mechanism of a hypervisor to emulate selected virtual cores at their proper speeds. Such an adjustment might violate timing constraints.

For example, an embedded system has two physical cores both operating at frequency *f*, where each physical core serves one virtual core in an initial configuration. Suppose the operating frequency of the first virtual core is adjusted to 0.6*f* and the operating frequency of the second virtual core is adjusted to 0.4*f*. After the frequency adjustments, the hypervisor might decide to turn off one physical core and to let the two virtual cores share one physical core for energy saving (Fig. 7). Suppose the two virtual cores use the first 6 μs and the last 4 μs respectively for every 10 μs time period over the physical core. This virtual core emulation might result in a deadline violation problem if the second virtual core runs a real-time task with a period of 5 μs and an execution time of 2 μs (at operating frequency *f*). There is a problem with emulating virtual cores over physical cores when timing constraints of processes must be satisfied.

Thus, two major design issues need to be addressed for DVS support of virtual cores: (1) How to model the DVS needs of a virtual core and (2) how to map the application requirements of virtual cores into the DVS settings of physical cores. Modeling the DVS needs of a virtual core should result in a modeled workload for the virtual core. A resource scheduling mechanism can then be proposed so that a hypervisor can schedule the workload execution.

A virtual core provides an environment for executing tasks in a similar way to a physical core with a user-specified frequency. To allocate proper execution cycles to tasks executed over virtual cores, each virtual core can be served by a Constant Bandwidth Server with Hard-Reservation (CBS-HR) [18],[19]. The CBS is scheduled and is designed to serve soft real-time tasks. It is also designed to serve hard real-time tasks in a system by using the Earliest-Deadline-First (EDF) scheduling algorithm—which always schedules a task with the earliest (absolute) deadline. Because CBS suffers from deadline aging, CBS-HR is an extension that guarantees a fixed execution budget *C* in every *T* time units, where *C* is the maximum budget (in terms of execution cycles) of a CBS-HR server, and *T* is the replenish period of a CBS-HR server.

In our virtualization system, the ith virtual core is modeled according to characteristics of the workload over it and according to three parameters (C_i, T_i, F_i). C_i is the maximum budget and T_i is the replenish period of the virtual core's corresponding CBS-HR server. F_i is the virtual core's user-set operating frequency. C_i is measured in cycles, T_i is measured in seconds, and F_i is measured in MHz. Setting C_i and T_i of the corresponding CBS-HR server to the maximum value permitted should redress any inequality. While there are some timing issues resulting from the workload over the virtual core, some additional constraints on setting the replenish period and maximum budget are incurred.

For multimedia applications and batch applications, users might be interested not only in the throughput of the virtual core but also in the response time of a certain workload. While the throughput is reflected by the operating frequency of the virtual core (set by the user), the replenish period of the virtual core affects whether reasonable response time delay is achievable. For example, a virtual core emulated on a physical core can be used to execute a certain workload (Fig. 8, lower), where the gray area is the execution cycles of the workload. Compared with the behavior when the same workload is executed on a physical core with the same emulated frequency (Fig. 8, upper), delay *D* in response time might be incurred. Therefore, a user of the virtual core might specify a tolerable delay σ that should not be exceeded, where response time delay in the virtual core with frequency F_v is defined as the difference between the response time of executing a workload in the virtual core and that of a physical core with the same frequency F_v, (*D* in Fig. 8). Let F_p be the frequency of the physical core. According to Fig. 8:

$$D = (T - \frac{C}{F_p}) + \frac{\Delta F_v}{F_p} - \Delta \qquad (1)$$

By substituting $X \cdot T$ for Δ, D becomes

$$T((X-1)(\frac{F_v}{F_p}-1)) \qquad (2)$$

which is a function of *x*.

Since F_v of the virtual core is not greater than F_p of the physical core, D has its maximum value when *x* approaches 0. That is, *D* is not greater than $T(1-\frac{F_v}{F_p})$ (3)

Combined with $T \leqslant \frac{\sigma}{1-\frac{F_v}{F_p}}$, D is not greater than the tolerable response time delay σ.

While the physical cores have DVS, the hypervisor of a power-aware virtualization system should be able to adjust the frequencies of physical cores for the purpose of saving energy. In our virtualization system, DVS in the physical core is exploited by integrating a DVS scheduling policy with the admission control mechanism. As a result, once our admission control mechanism is triggered, it invokes the DVS scheduling policy if DVS is supported in the physical cores. Given remaining utilization of a physical core, the DVS scheduling policy scales down the operating frequency of the physical core to the lowest available frequency so that remaining utilization is no less than 0. The scheduling policy only asks the admission control mechanism to reject the request if remaining utilization is less than 0, even if the physical core operates at the highest available frequency.

References

[1] P. Chun and B. Maniatis, "Augmented smartphone applications through clone cloud execution," in *Proc. 12th Workshop on Hot Topics in Operating Systems (HotOS XII)*, Monte Verità, Switzerland, 2009, p. 8.

[2] A. Rudenko, P. Reiher, G. Popek, and G. Kuenning, "Saving portable computer battery power through remote process execution," *SIGMOBILE Mobile Comput. and Commun. Review*, vol. 2, pp. 19-26, 1998.

[3] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A Berkeley view of cloud computing," EECS Dept, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28, Feb. 2009.

[4] S. Osman, D. Subhraveti, G. Su, and J. Nieh, "The design and implementation of Zap: A system for migrating computing environments," in *Proc.5th Symposium on Operating Systems Design and Implementation*, Boston, MA, 2002, pp.361-376.

[5] C. Keir, C. Clark, K. Fraser, S. H, J. G. Hansen, E. Jul, C. Limpach,I. Pratt, and A. Warfield, "Live migration of virtual machines," in *Proc. 2nd ACM/USENIX Symposium on Networked Systems Design and Implementation*, Boston, MA, 2005, pp. 273-286.

[6] Android Developers. [online]. Available: http://developer.android.com/

[7] R. Kouzes, J. Myers, W. Wulf, "Collaboratories: doing science on the Internet," *IEEE Computer*, vol.29, no.8, pp.40-46, Aug. 1996.

[8] J. Stone, D. Gohara, Guochun Shi, "OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems," *Computing in Science & Engineering*, vol.12, no.3, pp.66-73, May-June 2010.

[9] R. Nathuji and K. Schwan, "Virtualpower: coordinated power management in virtualized enterprise systems," in *Proc.21st ACM SIGOPS Symposium on Operating Systems Principles (SOSP'07)*, Stevenson, WA, 2007, pp. 265-278.

[10] J. Stoess, C. Lang, and F. Bellosa, "Energy management for hypervisor-based virtual machines," in *Proc. 2007 USENIX Annual Tech. Conf.*, Santa Clara, CA, pp.1-14.

[11] Y. Wang, X. Wang, M. Chen, and X. Zhu, "Power-efficient response time guarantees for virtualized enterprise servers," in *Proc. 2008 Real-Time Systems Symp. IEEE Comput. Soc.*, Barcelona, Spain, 2008, pp. 303-312.

[12] Y. C. Lin, C. Y. Yang, C. W. Chang, Y. H. Chang, T. W. Kuo, and C. S. Shih, "Energy-Efficient Mapping Technique for Virtual Cores," in *Proc. Euromicro Conf. on Real-Time Systems*, Brussels, Belgium, 2010, pp. 66-75.

[13] J. Flinn, D. Narayanan, and M. Satyanarayanan, "Self-tuned remote execution for pervasive computing," in *Proc. Hot Topics in Operating Systems (HotOS-VIII)*, Elmau, Germany, 2001, pp. 61-66.

[14] R. Balan, J. Flinn, M. Satyanarayanan, S. Sinnamohideen, and H. I. Yang, "The case for cyber foraging," in *Proc. 10th ACM SIGOPS European Workshop*, Sintra, Portugal, 2002, pp. 87-92.

[15] E. Cuervo, A. Balasubramanian, D. ki Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl, "MAUI: Making smartphones last longer with code offload," in *Proc. ACM MobiSys*, San Franisco, CA, 2010, pp. 49-62.

[16] M. Satyanarayanan, B. Gilbert, M. Toups, N. Tolia, A. Surie, D. R. O'Hallaron, A. Wolbach, J. Harkes, A. Perrig, D. J. Farber, M. A. Kozuch, C. J. Helfrich, P. Nath, and H. A. Lagar-Cavilla, "Pervasive personal computing in an internet suspend/resume system," *IEEE Internet Comput.*, vol. 11, no. 2, pp. 16-25, 2007.

[17] "100,000 Android Applications Submitted To Date, AndroLib Claims," techCrunch web site. [online] Availaible: http://techcrunch.com/2010/07/30/android-market-100000/

[18] A. Mancina, D. Faggioli, G. Lipari, J. N. Herder, B. Gras, and A. S. Tanenbaum, "Enhancing a dependable multiserver operating system with temporal protection via resource reservations," *Real-Time Syst.*, 2009, vol. 43, no. 2, pp. 177-210.

[19] L. Marzario, G. Lipari, P. Balbastre, and A. Crespo, "Iris: A new reclaiming algorithm for server-based real-time systems,"in *Proc. 10th IEEE Real-Time and Embedded Tech. and Applicat. Symp.*, Toronto, Canada, 2004, pp. 211-218.

Biographies

Shih-Hao Hung (hungsh@csie.ntu.edu.tw) joined the Department of Computer Science and Information Engineering at National Taiwan University as assistant professor in 2005. His research interests include cloud computing, parallel processing, embedded systems, and pervasive applications. He worked for the Performance and Availability Engineering group at Sun Microsystem Inc. in Menlo Park, California (2000-2005) after completing post doctoral work (1998-2000), Ph.D. (1994-1998) and M.S. (1992-1994) at the University of Michigan, Ann Arbor. He graduated from National Taiwan University with a BS degree in electrical engineering in 1989.

Tei-Wei Kuo (ktw@csie.ntu.edu.tw) is a Distinguished Professor in the Department of Computer Science and Information Engineering, National Taiwan University. He was the chairman of his department (2005-2008) and deputy dean of his college (2006-2008). Prof. Kuo is an IEEE Fellow. He is on the editorial board of IEEE Transactions on Industrial Informatics. He was the program chair and general chair of the IEEE Real-Time Systems Symposium in 2007 and 2008. He has published over 180 journal and conference papers.

Chi-Sheng Shih (cshih@csie.ntu.edu.tw) has been an associate professor at the Department of Computer Science and Information Engineering at National Taiwan University since 2008. In 2003, he received his Ph.D. from the University of Illinois at Urbana-Champaign. His research interests include real-time operating systems, real-time scheduling theory, embedded software, and software/hardware co-design for system-on-a-chip. He received the Best Student Paper Award from the IEEE Real-Time System Symposium in 2004 and Best Paper Award from the IEEE International Conference on Embedded and Real-Time Computing Systems and Applications in 2005.

Jeng-Peng Shieh (d97026@csie.ntu.edu.tw) received his B.S.E.E. and M.EMBA degrees from the National Cheng Kung University, Taiwan, in 1986 and 2007. He has been a Ph.D. student in Computer Science and Information Engineering of National Taiwan University since 2008. He was a system engineer at BDC Co., Hsinchu, Taiwan (1988-1991), a senior and chief system engineer at RPTI, Taipei (1991-1996), founder of a software house for designing Internet e-commerce system (1996-2002), the head of R&D department of PROTON Inc., Taiwan, (2003-2009). His research interests include linux-based mobile system design, power and performance optimization, system virtualization, and migration scheme leveraging clustering/cloud computing.

Chen-Pang Lee (d97039@csie.ntu.edu.tw) was born in Taipei, Taiwan. He received a B.S. degree in electrical engineering from Fu Jen Catholic University in 1996, and an M.S. degree in control engineering from National Taiwan University of Science and Technology in 1998. He is currently a Ph.D. candidate in information engineering in National Taiwan University. His research interests include wireless sensor networks and storage systems, especially for data synchronization and security.

Che-Wei Chang (f95093@csie.ntu.edu.tw) received a B.S. degree in Computer Science from National Chiao Tung University, Hsinchu, Taiwan, in 2006. He is currently working towards a Ph.D. in computer science and information engineering from National Taiwan University, supervised by Prof. Tei-Wei Kuo. His current research interests include energy-efficient scheduling, fast-booting designs, and multicore management.

Jei-Wen Wei (r98922062@csie.ntu.edu.tw) received a B.S. degree from National Chung-Cheng University in 2009. In 2009, she joined the Embedded Systems and Wireless Networking Laboratory under the supervision of Prof. Chi-Sheng Shih, in the Department of Computer Science and Information Engineering, National Taiwan University. Her research interests include embedded system design, real-time scheduling, and multiprocessor optimization. She is the project manager of the System Software Designs for Heterogeneous Embedded Multicore Processors project.

Building a Platform to Bridge Low End Mobile Phones and Cloud Computing Services

Fung Po Tso, Lin Cui, Lizhuo Zhang, and Weijia Jia
(City University of Hong Kong)

Abstract: Two waves of technology are dramatically changing daily life: cloud computing and mobile phones. New cloud computing services such as webmail and content rich data search have emerged. However, in order to use these services, a mobile phone must be able to run new applications and handle high network bandwidth. Worldwide, about 3.45 billion mobile phones are low end phones; they have low bandwidth and cannot run new applications. Because of this technology gap, most mobile users are unable to experience cloud computing services with their thumbs. In this paper, a novel platform, Thumb−in−Cloud, is proposed to bridge this gap. Thumb−in−Cloud consists of two subsystems: Thumb−Machine and Thumb−Gateways. Thumb−Machine is a virtual machine built into a low end phone to enable it to run new applications. Thumb−Gateways can tailor cloud computing services by reformatting and compressing the service to fit the phone's profile.

Keywords: mobile cloud computing; low end mobile phone; mobile OS; middleware

1 Introduction

Thumb−in−Cloud is a platform that delivers cloud computing services to low end mobile phones, and it has the potential to serve millions of people.

1.1 Motivations

A low end phone typically only has built−in applications and no means of installing or removing applications. Such devices usually have a 104 MHz ARM−7 processor and a lightweight Operating System (OS). In 2010, there were more than 5 billion mobile phones in use in the world [1]. Of these, about 3.45 billion were low end phones. Although shipment of high end smartphones increased rapidly in recent years, low end mobile phones still played a dominant role in the market [2], particularly in Asia and Africa.

Cloud computing has become more widespread in recent years due to vigorous promotion by Google, Amazon, Yahoo and others. Cloud computing involves moving services, computation, or data off−site to a centralized internal or external facility or to a contractor. By making data available in the cloud, it can be more easily and ubiquitously accessed, often at much lower cost. The value of the data itself may also increase as opportunities for collaboration, integration, and analysis are enhanced on a shared common platform [3]. A mobile cloud user need only put fingertip commands into the Internet for cloud services to "flow/push" onto the terminal screen.

However, because it is often not possible to install new applications on low end phones, many users cannot access powerful cloud computing services and storage. To bring these services to users, fundamental technology gaps should be filled:

(1) System gap: A low end mobile phone only has built−in software. This is burnt into the Read−Only Memory (ROM) during manufacturing to keep the price of the phone low. None of these built−in applications allow access to cloud computing services or even installation of new applications that would facilitate access. This means low end phones are extremely isolated. Also, their below average hardware is usually incapable of handling advanced services.

(2) Hardware gap: Cloud computing services are often in multimedia format. For low end phones, the display ability

The paper is partially supported by CityU Applied Research Grant (ARG) under Grant No. 9667033, Shenzhen Basic Research Grant under No. JC200903170456A, Shenzhen−HK Innovation Cycle Grant under No. ZYB200907080078A, RGC General Research Fund (GRF), HK SAR under Grant No. CityU 114609, CityU Applied R & D Centre (ARD (Ctr)) under Grant No. 9681001, and China NSF under Grant No. 61070222/F020802.

has been burned in and fixed. With very limited processing power, display processing cannot be handled by a low end phone.

(3) Energy gap: Low end phones have very limited battery life and do not meet the needs for the most cloud computing tasks, especially realtime data services.

(4) Communication gap: High bandwidth and a stable Internet connection are usually required for accessing cloud services. However, in low end mobile phones a stable and secure Internet connection cannot be guaranteed. Furthermore, mobile Internet access in some developing countries is still expensive for the general public.

1.2 Contributions

We have designed and implemented a novel and sophisticated Thumb-in-Cloud platform to overcome these challenges. Thumb-in-Cloud operates commercially in China to provide cloud services to low end mobile users. It includes two components: Thumb-Machine and Thumb-Gateways. Thumb-Machine is a virtual machine embedded in low end phones to enable them to execute and manage software applications. It is like a web browser that accepts user input, relays it to Thumb-Gateways, and displays results on the screen. Thumb-Gateways are aware of the network environment and phone profile so that cloud computing services can be tailored to suit each low end phone.

To the best of our knowledge, Thumb-in-Cloud is the first system to bring cloud services to low end phones by loosening their system software and hardware. This is a step in a different direction for mobile computing because research communities have tended to focus on high end rather than low end mobile devices.

2 Related Works

A cloud service has three distinct characteristics that differentiate it from traditional hosting: It is sold on demand, typically by the minute or the hour; it is elastic, so that a user can have as much or as little of a service as they want; and the service is fully managed by the provider [4]. Significant innovations in visualization and distributed computing, as well as improved access to high-speed Internet and a weak economy, have accelerated interest in cloud computing [5]. Since the service provider hosts both the application and the data, the end user can use the service from anywhere [6].

There are several well-known contributors to cloud computing [7]. Amazon Web Services is the largest public cloud provider. A major issue in cloud computing, especially with public clouds, is protection of user data. One concern is that providers themselves may have access to unencrypted customer data—whether it is on disk, in memory, or transmitted over the network [8]. To limit this exposure, many sources have recommended never giving providers access to unencrypted data or keys.

As data moves into the cloud, storage companies are taking advantage of virtualization and adding more memory to the data center. Storage virtualization can improve existing storage hardware and make provisioning easier, while adding memory to the data center can make accessing information faster [9]. Caching on the edge of the Internet is becoming a popular technique to improve the scalability and efficiency of clouds. Caches belonging to a cloud cooperate in order to serve misses and to maintain freshness of the cached document copies [10].

Today, there are already some good examples of mobile cloud computing applications, including mobile Gmail, Google Maps, and some navigation apps. However, the majority of applications today still perform most of the processing and store most of the data on the mobile device itself rather than in the cloud. A report from ABI research predicts the number of mobile cloud computing subscribers will total nearly one billion by 2014. There are two reasons why ABI believes cloud computing will become a force in the mobile world. First, the number of users that cloud technology has the power to reach is far more than the number of smartphone users alone. Second, all that is needed is access to the Internet for access to mobile applications.

3 Thumb-in-Cloud

Thumb-in-Cloud resides at the edge of a generic cloud to provide lightweight cloud services to isolated low end mobile phones or a bridge to popular cloud services such a online photo album. Thumb-in-Cloud provides text, image, sound and gaming services. In general, lightweight cloud services provided by Thumb-in-Cloud are satisfactory for the needs of low end mobile phone users. If they are not, Thumb-in-Cloud can relay service requests to a generic cloud service provider such as Google for further action.

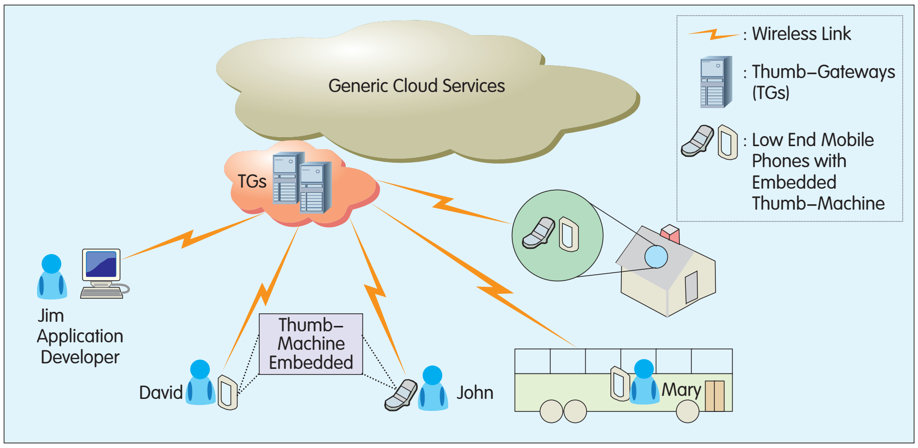
3.1 Overview

Fig. 1 is an overview of the Thumb-in-Cloud system. It comprises low end mobile phones and a group of supporting gateways. Implementing Thumb-in-Cloud is very challenging because low end phones are weak and every bit of them needs to be exploited to maximize processing capability. On the other hand, Thumb-Gateways need to carefully process content before pushing it to phones so that the content can be accepted.

In the following, we discuss four typical applications to show how Thumb-in-Cloud plays a critical role for cloud computing services on low end phones. Assume that Jim, David, John, and Mary have low end mobile phones. Jim is also an application developer (Fig. 1).

(1) Scenario 1: Jim is very satisfied with Thumb-in-Cloud. He has just developed a mini game and wants to release it to the public for free. He can do this by signing up to an application server in Thumb-Gateways and log his mini game to the application resource index.

(2) Scenario 2: David wants to receive digital entertainment but his low end phone does not allow it. With Thumb-Gateways and built-in



▲Figure 1. Architecture of Thumb–in–Cloud.

Thumb–Machine middleware he can download new games from Thumb–Gateways and install them on his handset with Thumb–Machine.

(3) Scenario 3: Mary has subscribed for weather information and daily news, and every morning content is pushed to her phone. She can check weather conditions before leaving for work. She can also read the daily news on the bus.

(4) Scenario 4: John likes taking pictures so he subscribed to a photo album service with a well–known cloud service provider. He can do this with his low end phone via Thumb–Gateways. He first downloads an application to connect to Thumb–Gateways, and then Thumb–Gateways relay his album access requests to the service provider.

3.2 Architecture

Fig. 2 shows Thumb–in–Cloud system components. A low end mobile phone with Thumb–Machine communicates with Thumb–Gateways via the General Packet Radio Service (GPRS) network. Thumb–Gateways have three operational groups for delivering content, for computation, or for security support.

We lodged a Thumb–Machine in target low end mobile phones to allow them to accept new applications. Thumb–Machine includes a component that manages installed applications and downloads new applications from cloud. It is also a virtual machine that can efficiently translate application scripts to machine code and execute them.

Thumb–in–Cloud cloud computing services are provided by Thumb–Gateways, and Thumb–Gateways are formed by a group of specific servers—an MP–aware functional group, a resource group, or a developer group—depending on their service nature.

3.3 Thumb–Machine

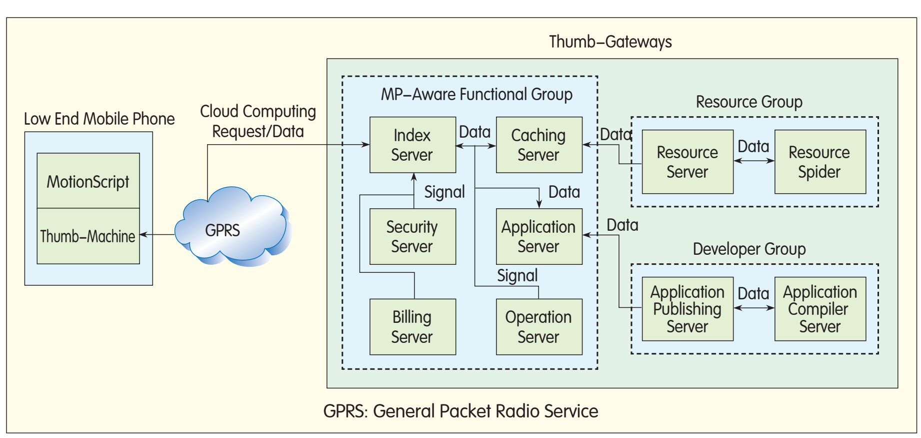
A high end phone can become programmable by downloading or embedding new application software whereas a low end phone has built–in applications that cannot be updated with new software installations. One of the objectives of Thumb–in–Cloud is to enable low end phones to be programmable using low bandwidth wireless channels through support of Thumb–Gateways system. By implementing a container–software called Thumb–Machine, low end phones can load new applications dynamically without the need for a user to intervene or even be aware of the process. The implementation is particularly focused on low end phones such as MTK phones with Nucleus OS that usually cannot satisfactorily access the cloud.

A typical low end phone from MTK costs US$30–50, which is reasonably affordable to the public. MTK phones often come with a 104 MHz ARM 7 processor and a lightweight OS.

High end phones achieve cloud computing access via a full operating system (iPhone OS, Andriod, Symbian, and Windows Mobile) and powerful hardware (1 GHz CPU, 512 MB RAM). Mid range phones usually have Java VM. Low end phones, which have the weakest hardware and lightest OS, often use binary loader or WAP to enhance their internet access capability. Although these solutions work well in their own classes, some are impossible to build into low end devices. Taking hardware limitations into account, we examined possible solutions to see if they fulfill user requirements. Comparisons of these possible solutions are listed in Table 1:

(1) Binary Loader: Although its object code is small and execution efficiency high, developing binary code is difficult for developers. In addition, binary code is highly dependent on platform because the binary level instruction set varies from platform to platform.

(2) WAP Browser: Wireless Application Protocol (WAP) is an open international standard for



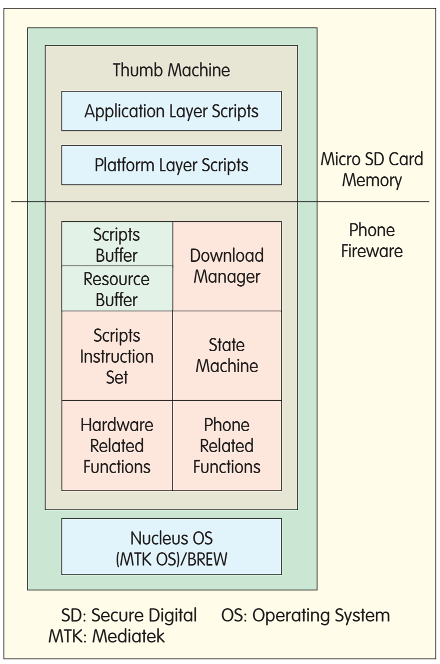
▲Figure 2. Thumb–in–Cloud architecture and signals.

D:\EMAG\2010–05–26/VOL8\COVER.VFT——5PPS/P28

▼Table 1. Comparison of binary loader, WAP browser, Java VM and Thumb–Machine

	Execution Efficiency	Application Portability	Development Difficulty	Application Object Size	Hardware Accessibility	User and Phone Interaction
Binary Loader	High	Hard	Difficult	Small	Easy	Good
WAP Browser	Low	Easy	Easy	Big	Difficult	Bad
Java VM	High	Easy	Easy	Big	Difficult	Fair
Thumb–Machine	High	Easy	Easy	Medium	Easy	Good

VM: Virtual Machine WAP: Wireless Application Protocol



▲Figure 3. Architecture of Thumb–Machine.

application–layer network communications. A WAP browser provides mobile users with web–like access to Wireless Markup Language (WML) websites. Text–based web applications always come up short in execution efficiency, hardware accessibility, and object size.

(3) Java VM: Java's philosophy is to let application developers "write once, run anywhere." Using a Java Virtual Machine (VM), Java's bytecode can be run without change. But for low end mobile phones, Java VM is too big for the OS, and porting Java VM to the MTK platform demands even more effort than developing new middleware.

These solutions have different strengths in execution efficacy (binary loader), development simplicity (WAP), or code portability (Java); but their weakness is that they cost too much to transplant into low end mobile phones. Binary loader lacks code portability; WAP is big and execution efficiency is low; and Java VM is too big and complicated to exist in MTK low end devices. Therefore, no existing alternatives are fit for low end mobile phones. Device weaknesses are detrimental to user experience. In light of this, a lightweight, scalable, efficient, and portable substitution is very beneficial to these devices.

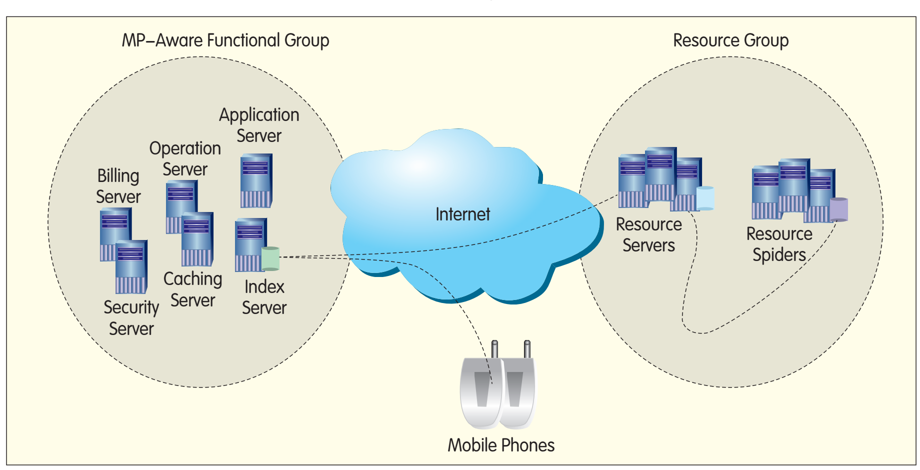
The architecture of Thumb–Machine is illustrated in Fig. 3. Together with the OS, Thumb–Machine should be burnt into the ROM as phone firmware. Thumb–Machine consists of seven components which coordinate to execute application scripts. There are two types of buffers residing in the Thumb–Machine: script buffer and resource buffer. When executing an application script, Thumb–Machine loads a target script (application) into script buffer. During execution, the script may ask for additional resources; for example, a picture to display or a sound to play. The resource files are loaded into the resource buffer beforehand in order to speed up execution efficiency. Thumb–Machine also has a download manager for downloading a resource that is not locally available on the fly. The idea behind downloading on the fly is downloading only what you need; thus GPRS traffic is reduced. Download Manager also monitors the whole download procedure to avoid potential errors triggered by improper abandonment of a download.

3.4 Thumb–Gateways

A Thumb–Gateway System is composed of several groups of servers (Fig. 4) aimed at bridging low end mobile phones with dynamic application services and content rich cloud computing.

Thumb–Gateways can be divided into three groups based on functionality: a mobile phone–aware functional group, a resource group, and a developer group. The operational group provides operation and computation support to the whole system; the resource group distributes and collects resources from the cloud that the user is interested in or has subscribed to; and the developer group enables developers to rapidly create and publish applications for mobile phones.

The mobile phone–aware functional group supports fundamental operations such as resource indexing, application provision, and data mining of content usage, history, and subscription. This group involves an index server, an



▲Figure 4. The Architecture of Thumb–Gateways system.

application server, an operation server, and sometimes a billing server. The index server is the entry point to Thumb-Gateways since every request arrives at the index server first. The index server interprets user requests and responds to them by dispatching the requests to actual destinations. The request can be classified as an application request or resource request. Index server redirects application requests to an application sever; and upon receiving a request, the application server pushes the application(s) to the user's mobile phone. If users request content, the index server queries the caching server first or passes the request(s) to the resource server if the caching server reports a negative result. The billing server charges users according to their subscriptions.

In this group, an operation server works in the background to analyze content consumption and subscription history so that user interest in specific content can be determined. The operation server also analyzes UI specifications and display format of mobile phones. The system is then able to adjust content so that it is provided in resource server in a way that best fits users' interests, mobile phone UI, and display. At the same time a fine tuned billing strategy ensures users are charged fairly and accurately.

In the index server, all packets returning to clients are compressed using zlib [11]—which has a compression ratio of around 2:1 to 5:1.

The resource group is composed of resource servers and resource spiders. The resource server mainly provides storage space for text, pictures, and multimedia files. Resource servers are distributed at the edge of the cloud, serving every corner of the Internet provided they are registered in the index server. As well as resource provision, resource servers also have a unique feature to alleviate computation burden of low end mobile phones. This feature is borrowed from cloud computing, which promotes thin clients by shifting intensive computation tasks to servers. In real implementation, every mobile phone service request carries a device fingerprint. Resource servers decode this fingerprint to extract the information such as screen resolution about a mobile phone and then pre-compute the proper organization of the content before it is pushed to the phone. As a result, a mobile phone need only display content on the screen without rearranging the content to be displayed.

In contrast, resource spiders are used for collecting content based on user subscription and analyzed results given by the operation server. Resource spiders retrieve desired information from the cloud regularly according to the information characteristics. For instance, resource spiders collect news provided free from news portals every morning at 5am. So users can read the news on the way to the office.

4 Conclusions

In this paper, we have presented an operational commercial system, Thumb-in-Cloud, for enabling cloud computing services access to low end mobile phones. In only one month, the system has been accessed by a wide range of low end mobile phone brands.

There are some technological hurdles in software, hardware, storage, and energy for some inexpensive mobile phones that prevent them accessing cloud computing. We have overcome these hurdles by introducing Thumb-Machine so that new applications (software) can be added onto the phone. Thumb-Gateways support cloud computing services or bridge them to general cloud computing services and Thumb-Gateways. We have elaborated on the design and implementation Thumb-Machine and highlighted some challenges encountered during research and development. Evaluation of Thumb-in-Cloud reveals that it fits very well into low end mobile phones and provides a more than satisfactory level of cloud computing service.

References

[1] BBC. [online]. Available: http://www.bbc.co.uk/news/10569081
[2] MobiThinking. [online]. Available: http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats
[3] M. Creeger, "Cloud computing: An overview," *ACM Queue-Distributed Computing*, vol. 7, no. 5, p. 2, Jun. 2009.
[4] M. A. Vouk, "Cloud computing Issues, Research, and Implementations," *J. Comput. and Inf.* Tech., vol. 4, pp. 235-246, 2008.
[5] B. Hayes, "Cloud computing," *Commun. ACM*, vol. 51, no. 7, pp: 9-11, 2008.
[6] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599-616, 2008.
[7] Infoworld. [online]. Available: http://www.infoworld.com/d/the-industry-standard
[8] A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Z. M. Armbrust, "Above the clouds: A Berkeley view of cloud computing," Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS 2009 28, Feb. 2009.
[9] S. Higginbotham,"Who will cache in on cloud storage?", [online]. Available: http://gigaom.com/2008/04/26/who-will-cache-in-on-cloud-storage/
[10] L. Liu and A. I. L. Ramaswamy, "Cache clouds: Cooperative caching of dynamic documents in edge networks," In *Proc. 25th Intl. Conf. Distributed Comput. Systems (ICDCS 2005)*, Columbus, OH, pp. 229-238.
[11] Zlib. [online]. Available: http://www.zlib.net

Biographies

Fung Po Tso (posco.cs@cityu.edu.hk) received B.Eng. and M.Phil. degrees from City University of Hong Kong. He is currently working towards a Ph.D. degree. His research interests include the design and applications of mobile communication systems, and system performance measurement and optimization.

Lin Cui (lincui2@student.cityu.edu.hk) is a Ph.D. candidate in the Department of Computer Science, City University of Hong Kong. He received his bachelor's degree from Shandong University in 2007, and a master's degree from Harbin Institute of Technology in 2009. His research interests include mobile computing systems and wireless ad hoc/sensor networks.

Lizhuo Zhang (lizhuocs@cityu.edu.hk) received his Ph.D. degree from the Central South University, China. He is currently a senior research associate in City University of Hong Kong. His research interests include multimedia communication, wireless networks, and mobile systems.

Weijia Jia (wei.jia@cityu.edu.hk) is a professor in the Department of Computer Science, City University of Hong Kong, and director of Future Networking Center, Shenzhen Research Institute of City University of Hong Kong. He received B.Sc. and M.Sc. degrees from Center South University, China, and a Master of Applied Science and Ph.D. from Polytechnic Faculty of Mons, Belgium. He was a research fellow at the German National Research Center for Information Science (GMD) in Bonn (St. Augustine) from 1993 to 1995. His research interests include next generation wireless communication, protocols and heterogeneous networks, distributed systems, and multicast and anycast QoS routing protocols. He has published in various IEEE journals, has contributed chapters to books, and has refereed international conference proceedings.

WiFace: A Secure GeoSocial Networking System Using Wi-Fi Based Multihop MANET

Lan Zhang[1], *Xuan Ding*[1], *Zhiguo Wan*[1], *Ming Gu*[1], *and Xiangyang Li*[2]
(1. Key Laboratory of Information System Security, Tsinghua University;
2. Department of Computer Science, Illinois Institute of Technology)

Abstract: A number of mobile Online Social Networking (OSN) services have appeared in the market in recent times. While most mobile systems benefit greatly from cloud services, centralized servers and communications infrastructure is not always available. Nor are location-based services offered to mobile devices without GPS. To take advantage of cloud and to address these problems, a Wi-Fi based multihop networking system called MoNet is proposed. On top of MoNET we propose a privacy-aware geosocial networking service called WiFace. Where there is no infrastructure, a distributed content sharing protocol significantly shortens the relay path, reduces conflicts, and improves data availability. Furthermore, a security mechanism is developed to protect privacy. Comprehensive experiments performed on MoNet show that the system is more than sufficient to support social networking and even audio and video applications.

Keywords: Wi-Fi; social network; privacy; MANET; WiFace

I Introduction

At the end of 2010, the number of mobile devices worldwide was about 5 billion. In conjunction with an Internet marketing company, we conducted a survey involving 215 respondents. The sample included people from a variety of professions with a wide range of educational backgrounds. 52% of respondents already had Wi-Fi interface, and 87% of the remainder had intentions of using a Wi-Fi enabled phone.

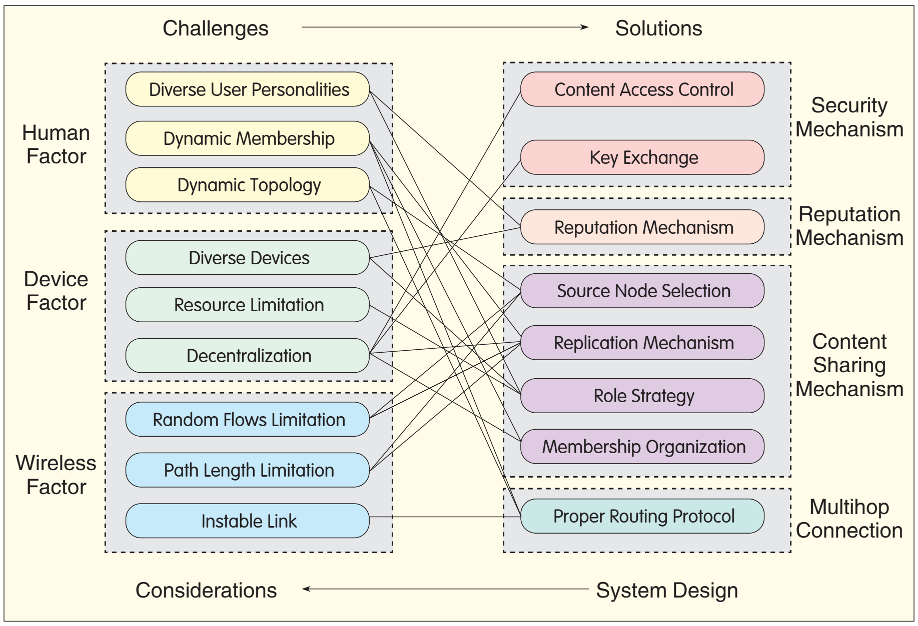
According to Schelling [1], [2], people who share mutual interests like to communicate and collaborate, and similar people tend to gather at the same places. Both our survey and [3] show that people in the same location tend to communicate and share information about their surroundings via mobile devices. Because these applications involve social relationships and sharing private information, 99% of respondents thought that privacy protection is necessary. In this paper, we focus on geographically co-located social networks, also called "geosocial" networks.

More and more Online Social Networking (OSN) sites are available on mobile phones. Typically, OSNs rely on centralized servers and the Internet, or on cellular networking access. This implies cost to end-users and access may not always be available for all users. Location techniques and devices are also required for these OSNs; and even with GPS, devices may still have problems with indoor use.

To meet application demands and address these problems, a secure geosocial networking system called WiFace is proposed. WiFace works when only some users have access to the cloud component, and it can also operate without any networking infrastructure or GPS module. In this paper, we address the problem of unavailable infrastructure for some or even all users. Multihop Mobile Ad hoc Network (MANET) [4] may be the best choice of network structure because it can be constructed easily using widely available mobile devices. Limited multihop coverage implies that end users are co-located. However, for many reasons, MANET is not practical for social networking applications without servers. Fig. 1 shows challenges on the left with solutions provided by WiFace on the right. These challenges include:

(1) Finding a suitable routing protocol: Due to dynamic topology and

This work is supported by National Natural Science Foundation of China under Grant No. 90818021, and 9071803.
The work of Li is supported by Tsinghua National Laboratory for Information Science and Technology(TNList), NSF CNS0832120, National Natural Science Foundation of China under Grant No. 60828003, National Basic Research Program of China ("973" Program) under grant No. 2010CB328100, and the National High Technology Research and Development Program of China ("863" Program) under grant No. 2007AA01Z180.



▲Figure 1. Challenges to design and implement a mobile geosocial networking system based on MANET (on the left), and our system solutions (on the right).

unstable link quality, it is difficult but crucial to find a suitable routing protocol that has been evaluated in real-life mobile scenarios.

(2) Limited capacity of MANET: Throughput and delay significantly decrease the number of hops in a MANET path [5], [6]. Gupta and Kumar [7] have shown that the asymptotic per-flow unicast capacity with n random flows of n nodes inside a unit square is $\theta(W/\sqrt{n \log n})$. These limitations require a carefully designed content sharing mechanism to shorten path length and reduce conflicts.

(3) Diverse devices and limited resources: Ordinary mobile devices usually have limited disk space, network bandwidth, and power. The condition of devices may also differ greatly. Thus full advantage must be made of resource-filled clouds and cooperation must be improved by role strategies.

(4) Dynamic membership and decentralization: Without a server or trusted third party, fast-changing membership and decentralization brings about great difficulty in managing membership, keeping content available, confirming identity, and authorizing access.

In the face of these challenges, we have designed and implemented a geosocial networking system consisting of a Wi-Fi based multihop network platform for personal mobile devices (called MoNet) and a distributed geosocial networking application with a security mechanism (called WiFace). This geosocial networking system works efficiently indoors and outdoors, with or without infrastructure or GPS module. The system was deployed in Tsinghua University for use by more than one hundred people, and comprehensive field experiments were conducted to evaluate its performance. The results show that MoNet can sufficiently support geosocial networking applications and even audio and video applications.

2 Application Description

As well as MoNet, two other applications were designed and implemented: WiFace and WiMarket. In this paper, WiFace is discussed in detail.

WiFace is a co-located social networking application with a security mechanism. It supports common social networking activities such as status updating, blogging, adding friends, and chatting. It also allows sharing of photos, videos, and documents and offers special co-located services. Take a mall for example. Stores can broadcast electronic product advertisements, hot sales, and discounts to customers in or around the mall. Customers can request or filter promotional information about a certain category of product or search surrounding stores. Interestingly, customers could also source others who wish to make a group purchase and perhaps get a better price. WiFace could also be used in a train, an international conference, an exhibition, or on a spring outing.

WiMarket is an online fleamarket application implemented on MoNet. One hundred and six heterogeneous mobile nodes including mobile phones and laptops were deployed in a 1200 m × 800 m area on campus. Experiments carried out on WiMarket are detailed in Section 4 of this paper.

Results show that MoNet has reasonably good network performance; thus, more audio and video applications can be implemented. MoNet can also be easily deployed and expanded greatly via the Internet and can be used as the basic network structure for wired range mobile communications—such as making a free transpacific call by mobile phone.

3 Architecture and Design

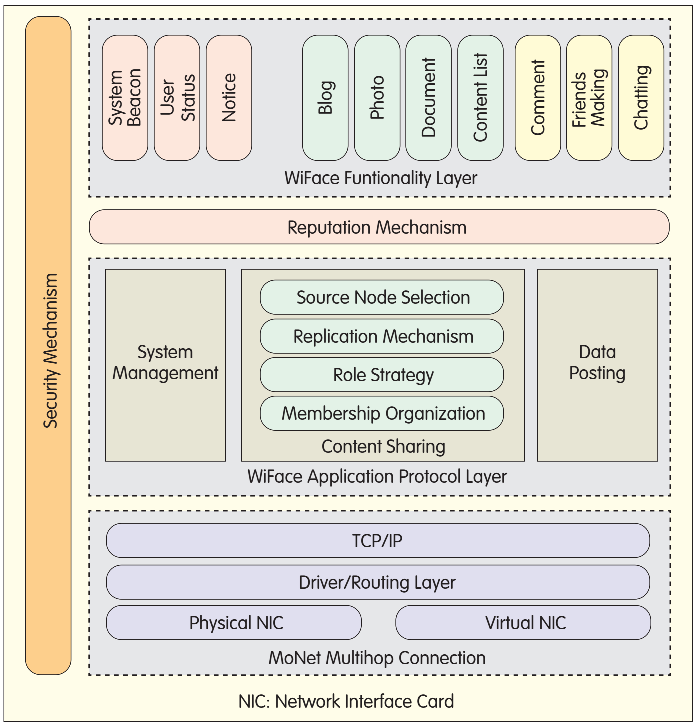
As illustrated in Fig. 2, the WiFace system consists of MoNet network platform and WiFace application built on top.

3.1 Network Formulation

The MoNet platform is based on a TCP/IP Reference Model and contains two primary layers:

(1) A Physical and Virtual Network Interface Card (NIC) Layer corresponding to the physical and data link layers. The virtual NIC is designed to support multihop communication as an additional network link without affecting other physical NICs.

(2) A Driver/Routing Layer located between the link and network layers that serves as a standard interface for them. This layer is the core of the



▲Figure 2. WiFace application structure.

MoNet platform. After MoNet has been installed, a virtual NIC is generated with a 48 bit virtual Ethernet address. Using the 48 bit virtual address for routing, the multihop routing protocol deals with frames from all network interfaces and sends them via the corresponding physical interface to the routing result. So the scope of MoNet can be expanded through physical links such as Internet, Bluetooth, and sensor. Mobile devices of MoNet can access the cloud as long as any one of them can connect to the servers.

3.1.1 Routing Protocol

Numerous routing protocols for MANET have been proposed and well studied. These include Ad hoc On-Demand Distance Vector (AODV) [8] and Dynamic Source Routing (DSR) [9]. A combination of link-state and DSR-style on demand querying was inspired by Mesh Connectivity Layer (MCL) [10]. However, most evaluations of these routing protocols are based on stationary nodes and simulations. To find a suitable routing protocol for MoNet, real mobile devices need to be evaluated and experiments carried out in typical scenarios. Performance is tested based on different link metrics—such as HOP, Round Trip Time (RTT), Expected Transmission Count (ETX) [11], and Weighted Cumulative Expected Transmission Time (WCETT) [12]. Since the per-hop RTT metric performs poorly due to self-interference [10], DSR is applied with three link quality metrics: HOP, ETX, and WCETT.

(1) Indoors

The main cause for reduced network capacity is collisions. Fig. 3 and Table 1 show indoor test results and, in this case, HOP performs the best. ETX performs almost the same as HOP but causes more probe overheads to the network and the metric calculation. Performance in an indoor mobile scenario was also evaluated. In this case, a sender carrying a mobile phone walks along a corridor to a receiver in the meeting hall at a speed of 1 m/s. Three other nodes are evenly placed along the corridor. Fig. 3 shows the maximum throughput of the TCP transfer between the sender and receiver during the 1 minute walk. The throughput of HOP increases by steps while ETX and WCETT have relatively smaller and more unstable throughput.

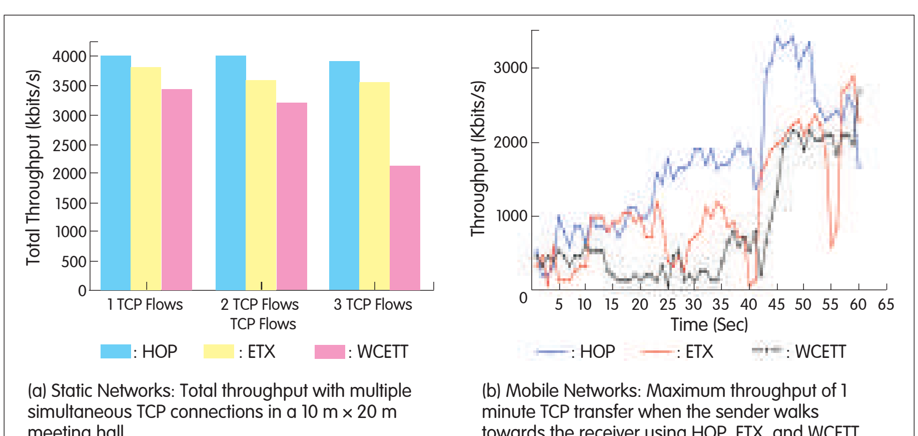
(2) Outdoors

Since users are usually moving around, node mobility is the main consideration. The test was conducted on a standard athletic field surrounded by a 400 m track. Four nodes were placed at the verteces of a square with sides 50 m located in the middle of the field. A sender and receiver on opposite sides of the track both move clockwise at the same speed of 1 m/s. The average throughput of HOP is 1.45 Mbit/s, ETX is 1.04 Mbit/s, and WCETT is only 316 kbit/s.

According to these results, the DSR protocol with HOP as the link metric is a more suitable choice for typical WiFace application.

3.1.2 Network Scope Expansion

The design of MoNet enables all mobile devices to access the cloud if any one of them has a connection with



(a) Static Networks: Total throughput with multiple simultaneous TCP connections in a 10 m × 20 m meeting hall.

(b) Mobile Networks: Maximum throughput of 1 minute TCP transfer when the sender walks towards the receiver using HOP, ETX, and WCETT.

ETX: Estimated Transmission Count WCETT: Weighted Cumulative Expected Transmission Time

▲Figure 3. Indoor test results based on different link metrics.

▼Table 1. Average path length and route change frequency with multiple 100 seconds simultaneous TCP connections in a 10 m × 20 m meeting hall

	Average Path Length (Hop)			Route Change Frequency (Time)		
	HOP	ETX	WCETT	HOP	ETX	WCETT
1 TCP Flow	1	1	1.46	0	0	2
2 TCP Flows	1.01	1.08	1.16	2	3	5.5
3 TCP Flows	1	1.14	1.64	2	2.25	22.5

ETX: Estimated Transmission Count WCETT: Weighted Cumulative Expected Transmission Time

the server. It also enables the scope of MoNet to be expanded through other physical links. Since most mobile devices do not have a globally unique IP address, a wired VPN can connect MoNets in different locations to greatly increase the scope (Fig. 4). In this way, the design and implementation of MoNet does not need to be modified, and no equipment is required other than an Internet computer with a globally unique IP address as the VPN server. This brings high expansibility to MoNet.

3.2 Application Design

The WiFace application layer is based on the connection of MoNet. System management, content sharing protocol, and data posting protocol are implemented to support respective upper layer functionalities (Fig. 2). Efficient lightweight security mechanisms across all layers are designed to protect user privacy.

3.2.1 Content Sharing Protocol

For mobile devices that cannot access the cloud, effective content sharing is crucial and challenging in a mobile geosocial networking system. This is due to highly dynamic topology and decentralization. Although systems like Gnutella [13] and FreeHaven [14] are purely decentralized, they are proposed for the Internet. In MANET, performance depends strongly on the path hop number. Recently, BitTorrent-like systems such as BitHoc [15] and BlueTorrent [16] have been adaped to wireless ad hoc networks. But they can suffer from single point failure of the initial seed, and the peer overlay is constructed on demand so that multihop capacity limitation and node mobility are not taken into consideration. Some algorithms of these systems are sophisticated but too complex for mobile devices with limited resources.

In WiFace, an active cooperative content sharing protocol is designed for mobile devices. To address the challenges in Section 1, this protocol takes advantage of short paths, resource-full devices (including servers when the cloud is accessible), and node mobility to offer efficient content sharing and good availability. By listening to other broadcast messages, every node maintains a member list (that records the behavior of all the other nodes) and a whole network content list. Queries are based on TTL limited flood, and a source node with a replica is chosen according to its distance and role level.

(1) Role Strategy and Reputation: A novel role strategy is proposed to improve cooperation among diverse devices by giving them different role levels. A node with more resources has a higher role level and automatically replicates more content items. When the cloud is accessible, a server holds the highest role level and replicates all content. The node with shorter hop path and higher role level is prioritized as the source node. To prevent a node from changing its role level, a reputation value is calculated based on its behavior recorded in other members' lists. A WiFace user spends his reputation points as currency on some functionalities and location-based services.

(2) Replication Mechanism: To keep content available in decentralized dynamic membership, a content item is not only stored by the creator but may also be replicated automatically by nodes with a role level greater than zero.

Both theoretical analysis and experiments show that the content sharing protocol can significantly decrease retrieval path length and prolong the content persistence time. Details about the content sharing protocol can be found in [17].

3.2.2 Security Mechanism

According to our survey, about half the respondents were worried about eavesdropping on private communications and/or content exposure to unauthorized people. In WiFace, different types of keys are used to protect private content and control content access.

(1) Friend Key Exchange: In WiFace, a pair of friends exchange a confidential symmetric key to form a secure channel for private communication and content sharing. Considering the limited computational capability of mobile devices, a lightweight scheme was designed

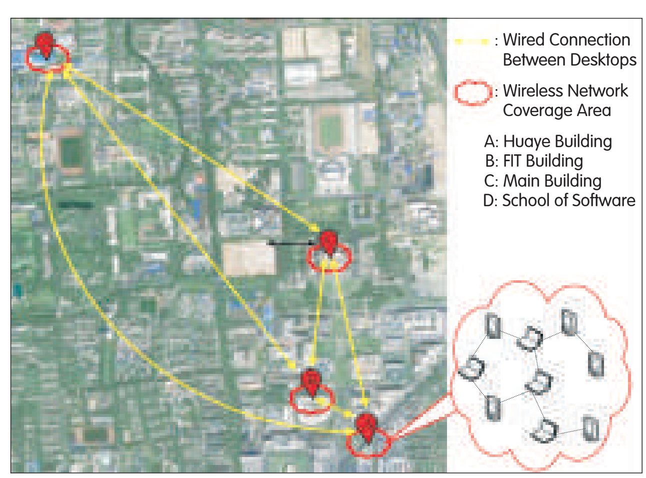


Figure 4.▶ MoNet experiment network topology with wired network connection. (Here, 106 nodes in 4 different buildings in Tsinghua university are connected using MoNet.)

▼Table 2. Maximum throughput and loss rate change with distance between two nodes

Distance	1 m	5 m	15 m
Throughput (Kbit/s)	4800	4680	2225
Loss Rate (%)	0.5	1.4	6.13
Distance (m)	30	40	45
Throughput (Kbit/s)	612	431	103
Loss Rate (%)	8.17	9.67	71.33

▼Table 3. Broadcast arrival rate

Node Number	10	20	50
Arrival Rate	99.8%	99.5%	96.9%

▼Table 4. Multihop maximum throughput and RTT

	1-Hop	2-Hop	3-Hop	4-Hop
Max Throughput (Kbit/s)	4318	2048	1142	323
RTT (ms)	18	49	105	340

RTT: Round Trip Time

based on shared knowledge. It combines two-party Elliptic Curve Diffie-Hellman (ECDH) [18],[19] and interlock protocol [20] to construct a friend key without a trusted authority. This can prevent Man-in-the-Middle (MITM) attacks to a certain extent.

The protocol for establishing keys between two individuals, Alice and Bob, works as follows:

- Alice sends a friend request message to Bob.
- Bob knows Alice in real life and receives the request. Bob then sends back an OK message to Alice; otherwise, he rejects the request (maybe sending a NO message).
- Alice and Bob generate the same confidential key k through ECDH. They formulate and send a friend challenge question to each other based on their shared private knowledge, which only the other should be able to answer.
- Alice encrypts her correct answer R_a with k and sends the first encrypted half to Bob. Bob waits to receive the first half of encrypted R_a then sends the first half of his encrypted answer R_b encrypted with k to Alice.
- Alice sends the other half to Bob until she receives Bob's first half.
- Bob decrypts the whole R_a and checks. If it is correct, he sends the other half of encrypted R_b to Alice, or else the exchange fails. Alice decrypts the whole R_b and checks. If it is correct, she sends an ACK, or else the exchange fails.

Only when they complete the whole protocol can they determine that the channel is secure. Then they bind the friend key k to each other's identity and communicate privately under the protection of k. If any step fails, the channel may have been compromised by an attacker.

(2) Content Access Control: After completing the process of making friends, each pair of friends shares a confidential 128 bit friend key, and every user has a friend key list. If a user wants to create friends-only content, a key is generated to encrypt the content and then the content key is encrypted and attached with the keys of authorized friends. In this way, the distributed authorization mechanism can be used to control content access efficiently.

4 Experiments and Evaluation

4.1 MoNet Performances

MoNet was evaluated based on 802.11b, which is supported by most mobile phone wireless interfaces.

(1) One-hop throughput and loss rates: Table 2 lists maximum throughput for one hop and loss rates according to the distance between two PDAs on a playground. The one-hop transmission range of PDAs is about 40 m.

Broadcast arrival rate: In this test, every node in a 10 m × 20 m room broadcasted 1 kb messages every second for 1000 seconds. Table 3 shows the broadcast arrival rate with different numbers of broadcasting nodes. The stable arrival rate means that system beacons and notices in WiFace are effective.

(2) Multihop performance: To test the multihop performance of MoNet, PDAs were arranged in a chain topology. As shown in Table 4, throughput decreases and latency increases significantly as path length increases.

Multiple simultaneous TCP connections: The total throughput of multiple simultaneous TCP connections was tested in a 10 m × 20 m meeting hall with 10 nodes, and the result is shown in Fig. 3. Even if there are 3 concurrent TCP flows in such a small area, the total throughput is about 4 Mbit/s with more than 1 Mbit/s for each connection. This shows that MoNet is capable of data stream applications like visual communication or video sharing.

4.2 Content Sharing Protocol of WiFace

This content sharing protocol was evaluated in two typical scenarios.

In scenario 1, there were 20 users in a playground. Three users published blogs at random times and the others read them randomly. This scenario shows that the content protocol can shorten the average path length from 3.1 hops to 2.3 hops without role strategy, saving 25% bandwidth and extra energy consumption. Furthermore, with role strategy the protocol can shorten the average path length from 2.9 hops to 0.7 hops, and even result in 0 hop path (requesting content from one's own replicas). This obviously enhances user experience.

In scenario 2, only one source node was used to publish a blog in the beginning. The result demonstrates that the content sharing protocol can significantly improve content persistence and availability in the changing population.

4.3 Real Usage Results

The WiMarket application on top of MoNet was also implemented, and real usage of MoNet was evaluated. Fig. 4 shows the experimental scenario; 106 heterogeneous mobile nodes were deployed in four buildings in Tsinghua University for the purposes of online trading. Four co-located MoNets in different buildings were connected by wired VPN connections. The average maximum throughput of a 3 hop path with a wired link between two nodes in different buildings is about 3.8 M. Eighty three trades were successfully completed within a 3 hour period.

5 Related Work

MoNet is a novel geosocial networking system that works efficiently with or without central servers or Internet access. It is also a wireless ad hoc network that is different from others. There have already been numerous attempts to address the problems of wireless ad hoc networks. These have contributed greatly to wireless ad hoc network research; however, most of them are indoor testbeds and not systems used in real life. This means they cannot be deployed on ordinary mobile phones and have less consideration of node mobility. There is still a lack of implementation and evaluation of wireless ad hoc networks for real use mobile devices. MoNet is a large-scale MANET composed of ordinary personal mobile devices and can be easily deployed indoors or outdoors to support real use applications. These experiments on MoNet can provide a new perspective on MANET.

6 Conclusions

In this paper, a secure geographical social networking system based on multihop MANET is discussed. Compared with existing social networking systems, this system can be easily constructed and can operate efficiently without relying on a central server or Internet access. Furthermore, it offers location-based services to low end mobile phones indoors or outdoors without the need for GPS modules. A distributed content sharing protocol with a role strategy and lightweight security mechanism deal with the challenges posed by a wireless connection, device constraints, and user behaviors. Privacy is also protected. The system was deployed in Tsinghua University and comprehensively evaluated. Results show that with a proper system architecture and well-designed protocols, social networking over MANET is feasible, and MoNet is even sufficient for audio and video applications.

References

[1] T. Schelling, "Models of segregation," *American Economic Review*, vol. 59, no. 2, pp. 488-493, May, 1969.
[2] T. Schelling, "Dynamic models of segregation," *Journal of Mathematical Sociology*, vol. 1, no.2, pp. 143-186, 1971.
[3] M. Matuszewski, N. Beijar, J. Lehtinen, and T. Hyyrylainen, "Understanding attitudes towards mobile peer-to-peer content sharing services," in *PORTABLE'07*, Orlando, FL, pp. 1-5.
[4] Mobile ad hoc networking (MANET): Routing protocol performance issues and evaluation considerations, IETF network working group RFC 2501, 1999.
[5] J. Li, C. Blake, D. De Couto, H. Lee, and R. Morris, "Capacity of ad hoc wireless networks," in *ACM MobiHoc 2001*, Long Beach, CA, pp. 61-69.
[6] X. Li, "Multicast capacity of wireless ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 950-961, Jun. 2008.
[7] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 22, no. 6, pp. 644-654, 1976.
[8] C. Perkins and E. Royer, "Ad-hoc on-demand distance vector routing," in *WMCSA 1999*, New Orleans, LA, pp. 90-100.
[9] D. Johnson and D. Maltz, "Dynamic source routing in ad hoc wireless networks," *Mobile Computing*, T. Imielinski and H. Korth, Eds., New York: Kulwer Academic Publishing, 1996, pp. 153-181.
[10] R. Draves, J. Padhye, and B. Zill, "Comparison of routing metrics for static multihop wireless networks," in *ACM SIGCOMM 2004*, Portland, OR, pp. 133-144.
[11] D. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A high throughput path metric for multihop wireless routing," in *ACM MobiCom 2003*, San Diego, CA, pp. 134-146.
[12] R. Draves, J. Padhye, and B. Zill, "Routing in multiradio multihop wireless mesh networks," in *ACM MobiCom 2004*, Philadelphia, PA, pp. 114-128.
[13] Gnutella. [online] Available: http://gnutella.wego.com
[14] R. Dingledine, M. Freedman and D. Molnar, "The Free haven project: Distributed anonymous storage service," in *Proc. of the Workshop on Design Issues in Anonymity and Unobservability*, Berkeley, CA, 2000. pp. 67-95.
[15] A. Krifa, M. Sbai, C. Barakat, and T. Turletti, "Bithoc: A content sharing application for wireless ad hoc networks," in *PerCom 2009*, Galverston, TX, pp. 1-3.
[16] S. Jung, U. Lee, A. Chang, D. Cho, and M. Gerla, "Bluetorrent: Cooperative content sharing for bluetooth users," In *PerCom 2007*, New York, NY, pp. 47-56.
[17] L. Zhang, X. Ding, Z. WAN, M. Gu, and Xiang Y. Li, "WiFace: a secure geosocial networking system using Wi-Fi based multihop MANET," in *Proc. 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond*, San Fransisco, CA, 2010, pp. 1-8.
[18] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 388-404, Mar. 2000.
[19] N. Koblitz, "Elliptic curve cryptosystems," *Mathematics of computation*, vol. 48, no. 177, pp. 203-209, 1987.
[20] R. Rivest and A. Shamir, "How to expose an eavesdropper," *Commun. of the ACM*, vol. 27, no. 4, pp. 393-395, 1984.

Biographies

Lan Zhang (zhanglan03@gmail.com) has a B.S. degree in computer software from Tsinghua University. She is currently a Ph.D. candidate at the Department of Computer Science and Technology, Tsinghua University. Her research interests include mobile ad hoc networks and mobile social networks.

Xuan Ding (dingx04@gmail.com) has a B.S. degree in computer software from Tsinghua University. He is currently a Ph.D. candidate at the Department of Computer Science and Technology, Tsinghua University. His research interests include mobile social networks, social network privacy and anonymity, and graph anonymity.

Dr. Xiangyang Li (xli@cs.iit.edu) is an associate professor of computer science at the Illinois Institute of Technology. He is also a visiting professor of Microsoft Research Asia from 2007 to 2008. He is recipient of China NSF Outstanding Overseas Young Researcher (B). The research of Dr. Li has been supported by USA NSF, HongKong RGC, and China NSF. His research interests span the wireless sensor networks, game theory, computational geometry, and cryptography and network security. He is a senior member of the IEEE and served various positions (as chairs and TPC members) at numerous international conferences.

Zhiguo Wan (wanzhiguo@mail.tsinghua.edu.cn) has B.S. degrees from the Departments of Automotive Engineering and School of Software, Tsinghua University. He received his Ph.D. degree in Computer Science from the National University of Singapore in 2006. From 2006 to 2008, he undertook postdoctoral work in the Computer Security and Industrial Cryptography research group, Catholic University of Louvain. He is currently a lecturer in the School of Software, Tsinghua University. His research interests include network security, security and privacy, and wireless security.

Ming Gu (guming@mail.tsinghua.edu.cn) has a B. S. degree in computer software from the National University of Defense Technology, China. She also has an M.A. degree in computer software from Graduate University of Chinese Academy of Sciences. She is currently the vice-director of the School of Software, Tsinghua University, and vice-director of Ministry of Education Key Laboratory of Information Security. Her research interests include software formal methods, trustworthy software, and middleware technology.

A Case for Cloud-Based Mobile Search

Yan Gao, Li Fu, Zhenwei Zhang, Shengmei Luo, and Ping Lu
(ZTE Corporation)

Abstract: Mobile search is beset with problems because of mobile terminal constraints and also because its characteristics are different from the traditional Internet search model. This paper analyzes cloud computing technologies—especially mass data storage, parallel computing, and virtualization—in an attempt to solve technical problems in mobile search. The broad prospects of cloud computing are also discussed.

Keywords: mobile search; cloud computing; parallel computing; virtualization

1 Introduction

The advantages of cloud computing as an IT infrastructure are becoming more apparent as Internet services develop. Cloud computing services follow a scalable delivery and usage model, which means services can be requested on demand and can be customized. Networks that provide resources are called clouds. Clouds form a kind of virtual data storage and resource pool on a web-based platform and these resources can be accessed by users.

There are three key cloud service models [1]: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). The core concept of cloud computing is to provide on-demand services and to make terminals into mere input/output devices by improving cloud processing capacity and reducing terminal processing load.

As an extension of Internet search technology, mobile search is a generic term for mobile-based search technology. It refers to the acquisition of on-demand information and services (Wireless Application Protocol (WAP) sites, Internet information, mobile value-added service, or local information) on a mobile terminal via Short Message Service (SMS), WAP, Interactive Voice Response (IVR) and other access modes [2].

2 Analysis of Mobile Search and Cloud Computing

2.1 Mobile Search Services Overview

Mobile search involves providing services via mobile search engines. These acquire information from user input and integrate information from different providers to build relationships between the two. Processing information on an engine means that information it can be made suitable for a mobile terminal. Compared with traditional Internet search, mobile search has the following advantages:

(1) More Flexibility: Users are no longer restricted by a fixed terminal and can search anywhere and anytime [3].

(2) Accurate Results: Mobile search is focused on simplicity and effectiveness. These features require the search engine to have stronger capacity for natural language analysis and to provide accurate results .

(3) Diverse Applications

Mobile search should be more like a basic capacity than a service. It has been widely used for entertainment data filing, travel and managing personal information [3]. Combining diverse features boosts the development of popular applications. Since the multiservice development model emerged, isolation of the mobile search services, and flexibility of search traffic have become clear requirements.

Many technical problems are encountered in the development of mobile search. These include:

(1) Data Expansion

For personalized search, more data from user access logs and other user-associated information needs to be recorded. These demand greater storage and processing capability in mobile phones.

(2) Limited Processing Power

Providing accurate results means increasing the processing workload of search technology. Enhancing user experience, processing search data, and correlation analysis all require quick and powerful computing.

(3) Information Security

For mobile search, the basic requirement of multiservice application is security. This is essentially different from Internet search technology.

(4) Service Flexibility

Service scale differs considerably during different stages of development. Fluctuations require powerful service processing capacity for a mobile search engine. As well as cost and energy saving, service flexibility must be taken into account.

2.2 Cloud Computing in Mobile Search

Key areas of mobile search development are mobile network, search technology, and user end. The emergence of cloud computing infuses a powerful driving force for the development of mobile search.

(1) Massive Data Storage

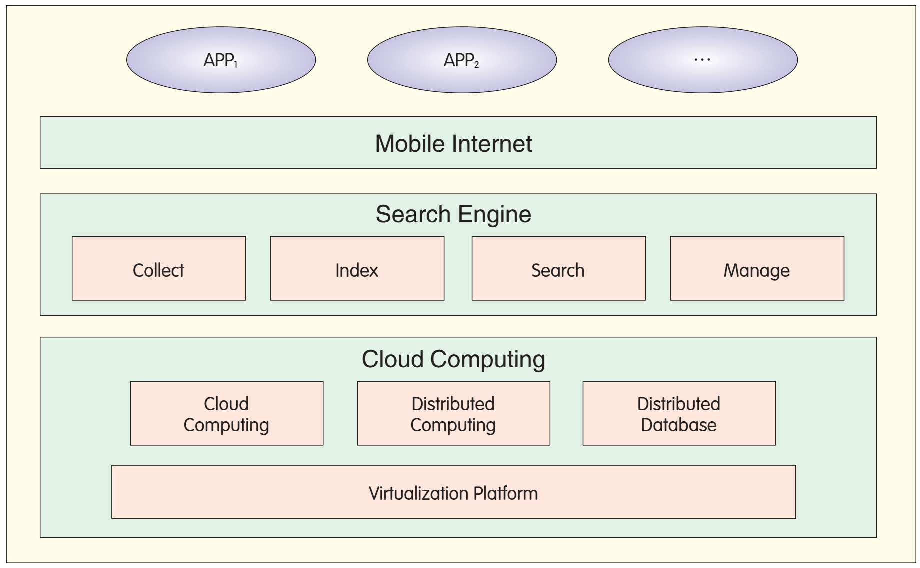
Cloud computing provides a secure and reliable data storage center for both storage and management. A cloud-based distributed network with scalable architecture not only provides mass data storage capacity but also takes advantage of storage capacity in the server itself. So computing and storage size are upgraded together, and this can significantly improve system reliability, availability, efficiency, and scalability.

(2) Parallel Computing

Parallel computing provides powerful computing for mobile search applications. MapReduce is a data processing solution based on mass data storage. Computing tasks can be decomposed by the MapReduce programming mode. Real-world computing tasks can be abstracted and then implemented by a partition statute using this programming model. Parallel computing is the solution to mass data storage in mobile search.

(3) New Solutions for Mobile Security

Cloud computing is combined with parallel computing, grid computing, and other emerging technologies and concepts. In applying cloud computing, security must be guaranteed for multiservice applications in the basic cloud structure. Distributed storage and distributed database are key technologies for which service security is a key requirement. Virtualization can be used to build a hardware and software wall for securing services [4].



▲Figure 1. System architecture of mobile search based on cloud computing.

(4) Flexible Platform

In cloud computing, mass resources are converged and integrated through virtualization in the resource layer, platform layer, and application layer, and through physical distributed integration. The cloud is not just a simple collection of resources. There is also a management mechanism so that the whole system can serve as a virtual pool for providing services and provide a flexible platform for multilevel applications of mobile search.

The cloud server generally processes complex calculations while mobile terminals are responsible for interaction with users. This reduces the processing workload in mobile devices and SaaS is implemented as well. Cloud computing also greatly reduces network requirement of mobile search [5].

3 Solution of Cloud-Based Mobile Search

In the following, a solution for cloud-based mobile search technology is introduced.

3.1 Architecture

In the architecture for the solution (Fig. 1), cloud computing provides storage computing capability and virtualization environment support. The core features—collection, index, search, and managing—of mobile search are designed and applied using cloud technologies. The system can be deployed as a basic search structure on a mobile Internet platform or as an independent mobile search service.

3.2 Key Technologies of Cloud Computing in the Solution

The key techologies for the solution include:

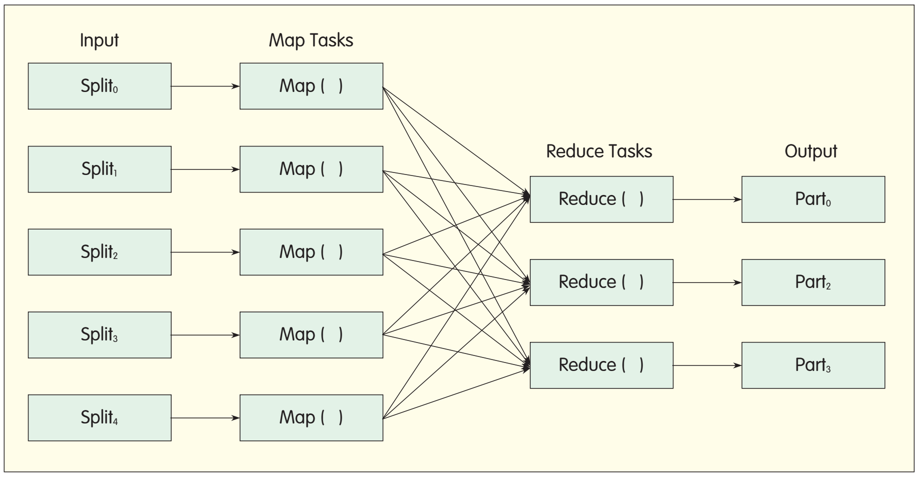
(1) Distributed Parallel Processing

Distributed computing and distributed storage are used to support mass data processing. Large-scale computing tasks can be divided and ruled by MapReduce (Fig. 2) to ensure system efficiency.

(2) Virtualization Platform

For multilevel applications of mobile search, a virtualization platform (Fig. 3) is introduced to provide a virtual service environment and to eliminate the incompatibility caused by different terminals. The virtualization platform supports the management and allocation of resources of heterogeneous virtual machines and provides a flexible virtual desktop with seamless interaction between servers and clients, seamless combination of applications, two-way audio and video transmission, and application deployment.

(3) An Organized Structure for Semantic Network and System of Text to Understand Chinese Webpages.



▲Figure 2. MapReduce distributed computing model.

Semantic understanding is the basis of understanding raw data and search purpose and is a key technology that makes mobile search results simple. We create a language model according to spoken Chinese and integrate sentence level text features based on spoken Chinese into the language model. At the same time, an unsupervised machine learning method recognizes named entities of page texts with unlimited fields to improve user experience. Semantic and syntactic decoding technology is also used to support natural language understanding.

3.3 Key Technologies of Mobile Search Based Cloud Computing

(1) Multiservice Index Sharing

Distributed file system, which distributes caching technology and virtualization technology, can isolate data in a multiservice environment. But search applications need to be provided with an index data sharing mechanism to improve search accuracy and data utilization. The core idea of this technology is as follows: Index data is placed into multiple sub-libraries to ensure data isolation; a raw service database is built on the distributed file system to ensure global shared access; index intermediate libraries are placed in the distribution database to ensure strict data isolation; and cross-service data access is controlled through a data dispatch server and data access layer to ensure only a service that has permission can read other service index intermediate libraries (Fig. 4). In specific implementations, this technology also further ensures data segregation and running effectiveness of services by accessing the image library when accessing other index libraries.

(2) Personal Real-Time Search

By combining a message-driven mechanism and distributed parallel computing architecture, a cloud-based search engine is realized. Such an engine has the capacity to update data in real time, which provides users with real-time search services. The platform informs the search engine to modify indexes automatically by message system when logging or modifying information or services. New search results are then returned immediately to ensure users can search the latest information the first time. A web-based search engine spends between one and three days updating search results. The key idea behind this capacity is to build an efficient and accurate distributed message-driven system. By classifying information and services, an efficient data distribution mechanism is built that balances data on computing nodes, and this ensures the message is processed efficiently and accurately. The messaging system is the second source for the search engine so that crawlers are not the only way to obtain information.

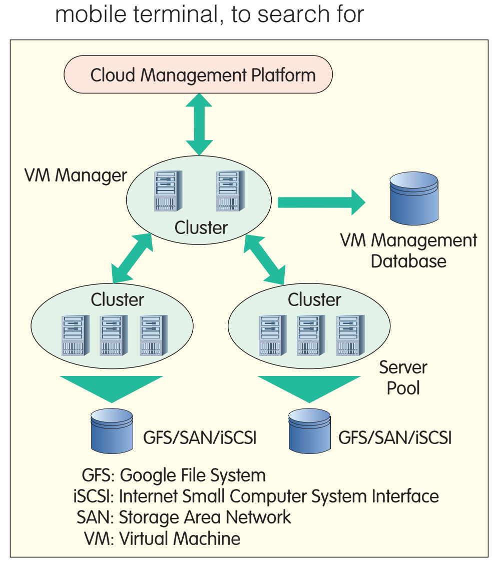
(3) Role Index

For the cloud computing mobile search platform, the user is introduced into the basic index and search process. Using this methodology, the search engine can predict the purpose of searches and provide a specific search "scene" that improves accuracy. An automatic association analysis mechanism analyzes the relationship between user role, user purpose, and category of actual search result. It then stores those relationships into a user's behavioral view. The search engine creates role indexes for information and services, and these bind the user roles. When a search call comes in, the search engine determines the most correlative role index according to the relationship stored in a user's behavioral view. It then starts an additional search on the role index in conjunction with the traditional search to improve accuracy.

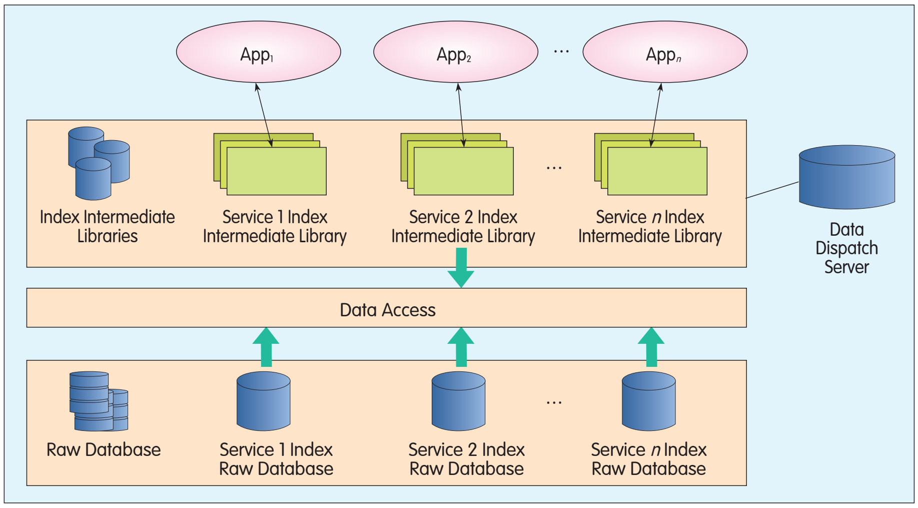
3.4 Application Case

The following illustrates the benefit of using cloud-based mobile search.

An enterprise has a set of news search systems for mobile users. The system servers provide the main search function and are combined with minicomputers with largest available 200 caps access traffic. There are still many redundant caps for traffic (average: 10 caps, max: 50 caps). A major incident happens, and people want to find out what is going on. So they use the most convenient tool, a mobile terminal, to search for



▲Figure 3. Virtualization platform deployment.



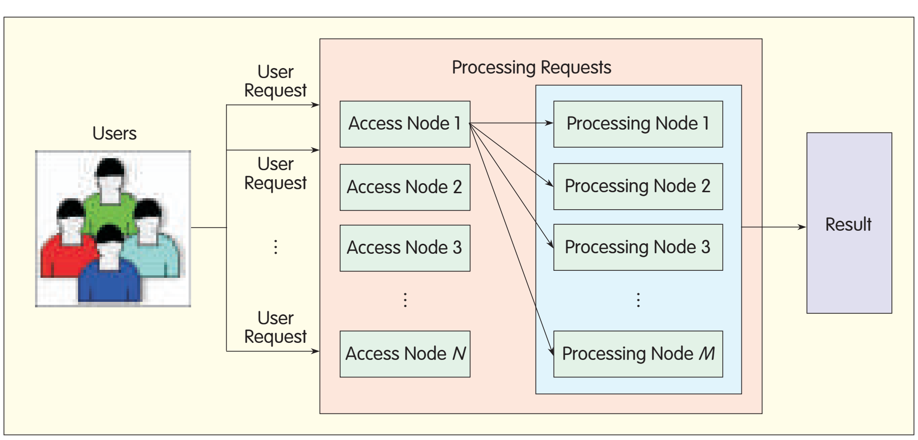
▲Figure 4. Multiservice index sharing structure.

information. This brings about a sharp rise in service access traffic to 200 caps. Such a spike in access traffic would challenge a mobile search system with a traditional structure. When the system is unable to complete user requests, a large number of requests accumulate in the system tray buffer, or are rejected, and this leads to delayed response times or denial of access. So user satisfaction with mobile search is greatly affected.

To solve this problem, the enterprise reconstructs the system based on cloud technology. Existing minicomputers are kept and are given the main computing processing capacity. When traffic rises sharply, the new search system with distributed structure and virtualization technology allows the new PC server to join into the service system (20-50 caps processing capability for one node) in one minute. With the added advantages of cloud computing (Fig. 5), the system assigns user requests to *N* access points, and the data of each access point is assigned *M* processing nodes for processing according to the computing capability of every node. In this way, large-scale service requests can be handled by multinode load sharing and multinode parallel processing. After the peak, the corresponding nodes can be released. In practice, one minicomputer and 6-8 PC servers doubles system capability to 400 caps a minute. Similarly, this can be used for support in emergency situations to ensure service availability.

4 Conclusion

As the basis of mobile Internet architecture, cloud computing will be vigorously developed. Technologies between mobile search and cloud computing will fuse together, and more services will be introduced with this converged technology that will bring about greater convenience for work and life in the twenty first century.



▲Figure 5. Processing flow for the user request.

References

[1] Y. Liu, "Summary of cloud computing and mobile cloud computing application," *Information and Communications Technologies*, No. 2, pp.14-20, Feb. 2010.

[2] Q. Yuan, "Mobile Search Technologies and Service Development," *Telecom Network Technology*, no. 4, pp. 38-42, Apr. 2007.

[3] T. Guo, "Competition Promoting Development: Mobile Search Trends," Globrand. [online]. Available: http://www.globrand.com/2010/344424.shtml

[4] Deyi Li, "Analysis of Hot Topics in Cloud Computing," *ZTE Communications*, vol. 8, no. 4, pp. 1-5, Dec. 2010.

[5] P. Zhao, "Cloud Computing Technology and Its Applications," *ZTE Communications*, vol. 8, no. 4, pp. 34-38, Dec. 2010.

Biographies

Yan Gao (gao.yan6@zte.com.cn) graduated from Chongqing University in 2010, majoring in Communication and Electronic Information. She is now a pre-research engineer with ZTE Corporation. Her research areas are Chinese word segmentation technology, video search technologies, and the relevant aspects of cloud computing technology.

Li Fu (fu.li3@zte.com.cn) is product and project manager at the Chongqing Institute of ZTE Corporation. His reasearch areas include cloud computing and the construction of product platforms.

Zhenwei Zhang (Zhang.Zhenwei@zte.com.cn) is head of the Chongqing Institute of ZTE Corporation. He is responsible for product operation and daily affairs of the Institute. His main research interest is relevant aspects of cloud computing technology and technologies of the Internet of Things.

Shengmei Luo (luo.shengmei@zte.com.cn) graduated from Harbin Institute of Technology in 1996, majoring in Communications and Electronic Information. He is a chief engineer and architect at ZTE Corporation. He is also a member of the China Cloud Computing Committee and heads pre-research into new technologies. He was awarded a second prize for scientific and technological progress with several invention patents. He has published a number of academic papers in core national communications journals.

Ping Lu (lu.ping@zte.com.cn) joined ZTE Corporation in 1996. He is the president of the Business Research Institute of ZTE Corporation. He has served as the minister of the Network Department, product manager of operational support, product manager of WDSS, and deputy general manager of business software products.

An On-Demand Security Mechanism for Cloud-Based Telecommunications Services

Zhaoji Lin, Ping Lu, Shengmei Luo, Feng Gao, and Jianyong Chen
(ZTE Corporation)

Abstract: As cloud computing gains in popularity, data migrated off premises is exposed to more threats than ever before. This is because data is out of control of the owner while floating in the cloud. Traditional device-centric security systems are not efficient enough and need to be evolved to data-centric protection systems. Cloud telecommunications services require security measures in three domains: data storage, processing, and transmission. Data stored in the cloud requires a mechanism to protect it; data in transit needs to be protected either at the service or transmission level; and data being processed needs to be protected during the processing stage. In this paper, we propose a security model based on a new method of security domain division to provide on-demand, dynamic, and differentiated protection for cloud-based telecommunications services.

Keywords: cloud computing; security; on demand

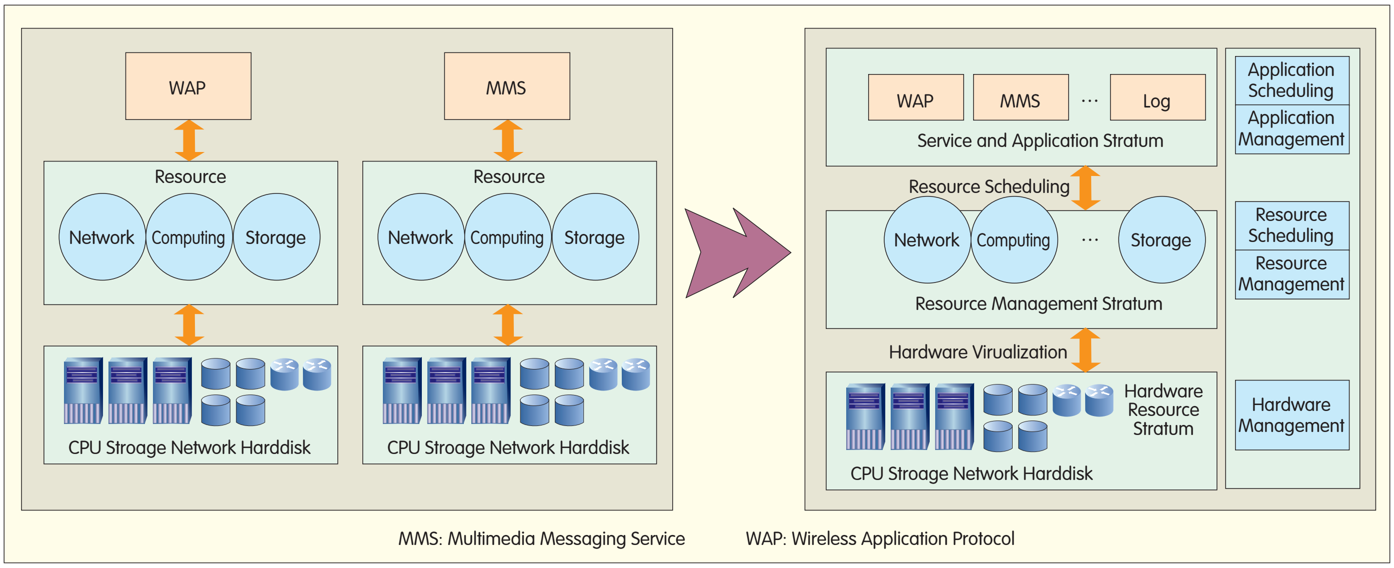
1 Introduction

With innovation in cloud technologies, services need to be rebuilt for the cloud. Commercial cloud applications include Amazon EC2/S3 [1], Google Apps [2], and Force.com [3]. In addition, Microsoft and Chinese carriers such as China Mobile, China Telecom, and China Unicom have also launched cloud services. Government activities are fast catching up to commercial activities; in the U.S. there is Apps.gov [4], the U.K. government operates G-Cloud, and the Canadian government also has a cloud.

Generally speaking, a cloud is discussed in terms of services, and services are being enriched and reinvented as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Cloud computing is a promising paradigm that has drawn attention from both academia and industry. By combining existing and emerging techniques from Service-Oriented Architectures (SOA) and virtualization, cloud computing resources in the computing infrastructure can be provided as services over the Internet. As promising as this sounds, cloud computing also faces many challenges that, if not resolved well, may impede its fast growth. Data security is of significant concern for users who store their sensitive information on cloud servers. These concerns are exacerbated by the fact that cloud servers are usually owned by commercial providers and are very likely to be outside the trusted domain of users. Data confidentiality in cloud servers is highly desired when data storage is outsourced. In some practical applications, data confidentiality is not only a security and privacy issue, but also a legal concern. A cloud is distinguished from other environments in that users may feel a vague insecurity about participating in a cloud, and this feeling cannot be easily overcome. In a public cloud, users can delegate system administration to cloud providers, but this also means that administration and operations are not controlled by the user. Furthermore, because of virtualization in multitenant services, there may be additional concerns about the physical proximity of data to competitors and protection of that data from competitors in a virtual environment. An IDC survey on Cloud/On-Demand showed that more than seventy percent of potential cloud users view security as a major reason against adopting clouds.

Cloud security has many facets, and researchers have discussed cloud security from their own viewpoints. Many of these researchers work in the cloud security alliance [5] and are making efforts to publish guidelines on security. Here, a dynamic, on-demand security mechanism is proposed to protect data and infrastructure in the cloud.

Domain division and a dynamic



▲Figure 1. Evolution of telecommunication service provision from silo to cloud-based.

▼Table 1. Comparison of traditional service environment security and CTSE security

Traditional Service Security Issues	CTSE Security Issues	Solving Technology
Individual Service Oriented	Total Services Oriented	On-Demand and Differentiated Security Mechanism
Affect Only an Individual Service While Be Attacked	Affect all Services While Be Attacked	Leveled Security Mechanism

CTSE: Cloud-Based Telecommunications Service Environment

on-demand/security mechanism can protect the data and infrastructure residing in security domains. Enhanced security can accelerate the deployment of cloud-based telecommunications services.

2 Cloud-Based Telecommunications Service Environment

In cloud computing, the traditional telecommunications service environment constructed in silo manner is transformed into the environment in a resource sharing model (Fig. 1). This transformation significantly decreases investment in service deployment and expansion.

The security requirements of Cloud-Based Telecommunications Service Environment (CTSE) are significantly different from those of a traditional service environment. A comparison of these two environments is given in Table 1. The technology used to solve CTSE security issues in column 2 is suggested in column 3 of Table 1.

3 A New Methodology for Studying Security in Cloud-Based Telecommunications Services

Data being migrated off premises is exposed to more threats than ever before because it is not within the reach and control of the owner while floating in the cloud.

In this regard, traditional device-centric security systems are being evolved to data-centric security systems. Cloud-based telecommunications service needs to be protected in three key domains: data storage, processing, and transmission, as shown in Fig. 2. The service must provide a mechanism to protect the stored data in the cloud, and data in transit needs to be protected either at the service or the transmission level. In most services, transmission level protection is chosen, and Secure Sockets Layer (SSL)/Transport Layer Security (TLS) protocols are used. Data also needs to be protected in the processing stage.

The relationship between data storage, processing, and transmission is shown Fig. 3.

The data storage and processing domains are connected with the end user through the data transmission domain. During its lifecycle, information can move through the data storage domain and data processing domain, and to the end user via the data transmission domain. At any moment, it

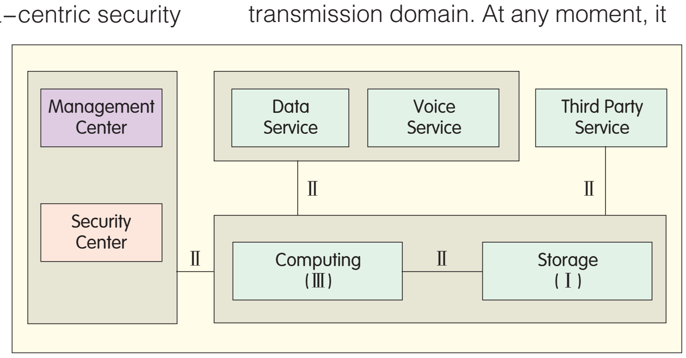
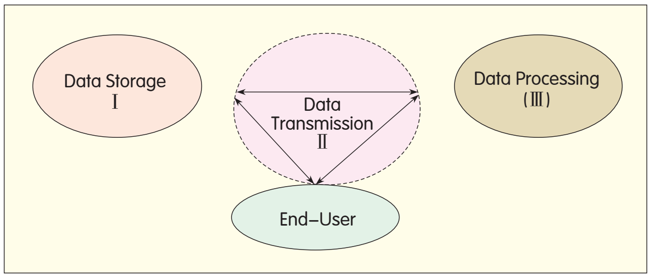


Figure 2. ▶ Security framework for cloud-based telecommunications service.



◀Figure 3. Relationship of data storage, processing, and transmission domains.

must be protected by the domain it resides in.

4 On-Demand Security Conceptual Model

A conceptual security model (Fig. 4) is given to illustrate how on-demand security can be achieved in a cloud-based telecommunications service environment.

Assume that vector $\vec{A}$ is the security unit set. Let $\vec{A}_1$ be the security unit set of the transmission security domain, $\vec{A}_2$ the security unit set of the processing security domain and $\vec{A}_3$ the security unit set of the storage security domain. $\vec{A}_1$, $\vec{A}_2$, and $\vec{A}_3$ are the subsets of $\vec{A}$; that is, $\vec{A}_1 \subset \vec{A}$, $\vec{A}_2 \subset \vec{A}$, and $\vec{A}_3 \subset \vec{A}$. Let $\vec{f}_1$, $\vec{f}_2$, and $\vec{f}_3$ be the security parameter vectors of the security unit sets of the above three domains. The expression of these security parameter vectors is $\vec{f}_i(x_1, x_2, x_3)$, where x_1 is service type, x_2 is site where the data is located, and x_3 is the level of security.

On-demand security for cloud-based telecommunications service is computed as $\vec{A}_1 \times \vec{f}_1 \oplus \vec{A}_2 \times \vec{f}_2 \oplus \vec{A}_3 \times \vec{f}_3$; that is, the integration of the security solutions of three domains, where $\oplus$ is the connector.

The security units in $\vec{A}_i$ are security functions such as encryption, authentication, and integration that were already realized by the cloud platform during the R&D stage.

$\vec{f}_i$ is the security assessment model that should be implemented by Security Operation Center (SOC). It is a mathematics model plus necessary security policies. In this model, a security administrator needs to configure parameters for x_1 and x_2, and the user configures the parameter for x_3. These parameters are only relevant to the service platform. Once a service and cloud platform has been decided, these parameters are determined accordingly.

The benefits of this new approach are:

(1) Each security domain faces the same type of security threats, which means the same security unit set is needed ($\vec{A}_i$ is the same). Division of security domains with the same security unit set is beneficial to SOC in establishing a corresponding policy (security parameter) model.

(2) For the service cloud platform, data transmission, processing, and storage is not necessarily provided by an individual operator. They could be provided by different operators or partly belong to a private cloud provider.

Security domain division proposed in this paper has at least four overall advantages:

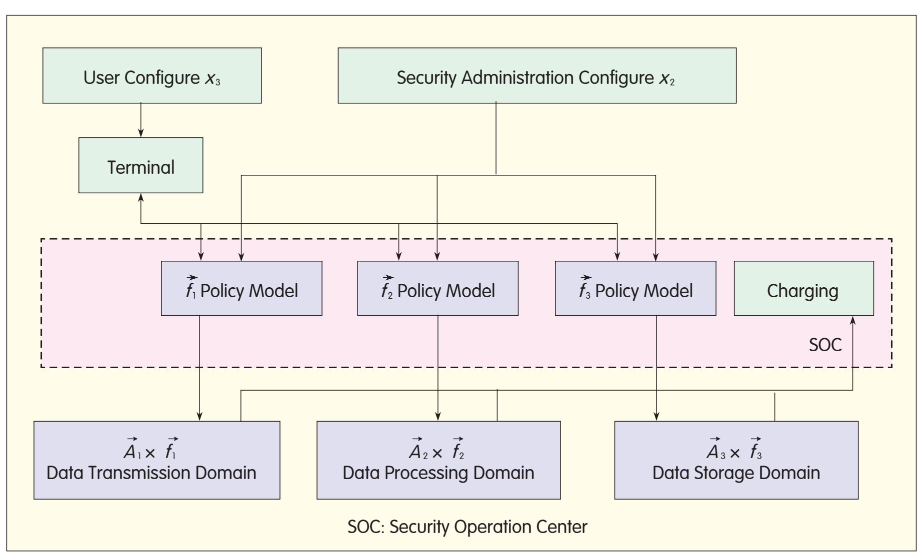
(1) Only 3 types of parameters need to be configured. Users configure the security requirements for a service (x_3), and SOC configures service location (x_2) and service type (x_1). Once a cloud platform is built, x_2 and x_1 are determined and kept static, while x_3 is determined by user requirements. So input of parameters is manageable and configurable.

(2) Each security domain constructs its own policy model depending on its own security unit characteristics. Changes of f_1 are independent from f_2 and f_3; that is, f_1, f_2, and f_3 are thoroughly decoupled. This simplifies the policy models and means they can be easily and accurately implemented.

(3) Only implementation of security unit technologies needs to be considered while execution is delegated to the policy model. In this way, the development of security modules involved is made easier. Existing network, services, and storage can be used simply by adding a configuration interface open to the policy model.

(4) The execution result can be fed back to a charging center so that on-demand security or Security as a Service can be provided.

Fig. 5 illustrates the application of this model in a real cloud-based telecommunications service scenario. Alice, Bob, and George are employees at the same company. Alice is staying in a hotel during a business trip, and Bob and George are both in the office. Alice initiates a video call with Bob and a text conversation with George to discuss the market strategy for next year (which requires high level



▲Figure 4. An on-demand security mechanism for cloud-based telecommunications services.

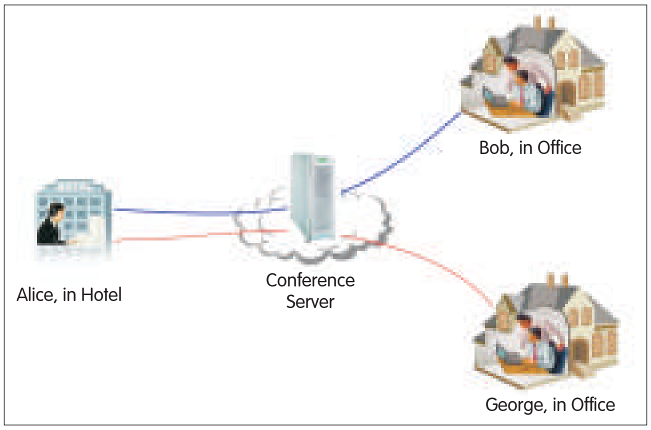


Figure 5. On–demand, differentiated security service for cloud–based conferencing.

confidentiality). The cloud–based conference server chooses the appropriate security mechanism for Alice, Bob, and George by acquiring the location indicator (can be determined from IP address), service type indicator, and the security assurance level that Alice, George and Bob set beforehand. The conference sever chooses a stronger authentication mechanism for Alice because she is in a less secure environment than Bob and George. For Alice, authentication using a password and usb key is necessary; while for Bob and George, only a password is required. Considering the confidential nature of their meeting and the high security assurance level selected by all of them, at least 256 bits or stronger encryption key and Advanced Encryption Standard (AES) encryption algorithm is needed. Data integrity protection is not applicable for communication between Alice and Bob so that high real–time performance can be achieved, but it should be applied between Bob and George to avoid the texts being tampered with. The security unit set in this case is the security capabilities supported by the cloud–based conference system.

5 Conclusion

In a cloud computing environment, on–demand and differentiated security services are of utmost concern to end users. We propose a new method using security domain division and present a conceptual security model based on these domains. This model can be used to provide dynamic, on–demand, and differentiated protection for cloud–based telecommunications. In this way, SECurity as a Service (SECaaS) can become reality at only small cost.

References

[1] AmazonWebServices, Amazon. [online]. Available: http://aws.amazon.com/

[2]GoogleApps, Google. [online]. Available: http://www.google.com/apps/intl/en/business/index.html

[3] Force.com, Salesforce.com. [online]. Available: http://www.salesforce.com/platform/

[4] http://www.info.apps.gov/[5]CloudSecurityAlliance. [online]. Available: http://www.cloudsecurityalliance.org/

Biographies

Zhaoji Lin (lin.zhaoji@zte.com.cn) graduated from Huazhong University of Science and Technology, majoring in Systems Engineering. He is a project manager for service and application security in the Standards Development and Industry Relations department of ZTE Corporation. His research interests include cloud computing and ubiquitous network and identity management. He has published more than 60 contributions and proposals on cloud computing security, identity management, device management, and digital rights management for ITU–T SG17, Open Mobile Alliance (OMA) and China Communications Standards Association (CCSA). He has also been an editor and convenor for several work items in these SDOs.

Ping Lu (lu.ping@zte.com.cn) graduated from South East University, China majoring in Automatic Control Theory and Application. He is chief executive officer of the Service Institute of ZTE Corporation. For more than a decade, he has steered the Institute towards innovative research and development of value added services, cloud computing, Internet services, ICT services, and home network services.

Shengmei Luo (luo.shengmei@zte.com.cn) graduated from Harbin Institute of Technology, China, majoring in Communications and Electronic Information. He is a chief engineer and architect at ZTE Corporation. He is also a member of the China Cloud Computing Committee, and heads pre–research into new technologies. He was awarded the second prize of scientific and technological progress, with several invention patents. He has published a number of academic papers in core national communication journals.

Feng Gao (gao.feng1@zte.com.cn) graduated from Beijing University of Aeronautics & Astronautics, majoring in Computer Science. He is a senior engineer in the Standards Development and Industry Relations department of ZTE Corporation. His research interests include bearer network controlling and network security. He has published more than 10 academic papers.

Jianyong Chen (chen.jianyong@zte.com.cn) received a Ph.D. degree from City University of Hong Kong in 2003. Currently, he is technical advisor at ZTE Corporation. His research interests include security of cloud computing and identity management. He has published more than 20 contributions and proposals on cloud computing security and identity management for the ITU–T SG17 and China Communications Standards Association (CCSA). He is also the vice–chairman of ITU–T SG17, chairman of ITU–T WP3/SG17 and chairman of fix network security working group in CCSA.

Multifrequency Networking Solution for TD-SCDMA

Abstract: This paper introduces the characteristics of TD-SCDMA, and analyzes some networking schemes and methods of multifrequency. For the 5 MHz frequency bandwidth, a frequency planning scheme containing three frequencies is examined, and a simulation model is built to validate the performance of this scheme. Finally, this paper analyzes the advantages and disadvantages of the scheme, and proposes some directions for the future study of networking planning.

Keywords: TD-SCDMA; multifrequency networking; simulation; frequency planning; multi-frequency

Min Jin
Wenbo Wang
Mugen Peng
(Key Lab of National Education Ministry for Pan-Network Wireless Communication, Beijing University of Posts and Telecommunications)

In radio communications, system capacity and interference directly affect user experience. Interference limits system capacity, so the primary task of radio network planning and optimization is to solve this contrary relationship. Users are also increasingly aware of networking technology within the system. How to provide a reasonable networking scheme, and increase system capacity and performance to the largest possible extent by reducing interference has become a recent focus of study.

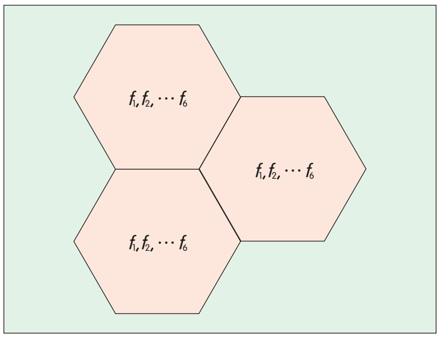
1 Multifrequency Networking Mode

Time Division-Synchronous Code Division Multiple Access (TD-SCDMA) is a 3G standard being pursued in China. It is a multiaccess technology and a hybrid of Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA) and Space Division Multiple Access (SDMA). TD-SCDMA adopts smart antennas, joint detection technology, and Time Division Duplex (TDD) mode. Symmetric frequency band is not required, but high frequency spectrum utilization and asymmetric data service are supported. TD-SCDMA carrier bandwidth is 1.6 MHz, and compared with 5 MHz Wideband Code Division Multiple Access (WCDMA), three carriers can be provided within the same bandwidth. Therefore, frequency planning is less complicated than in other 3G networks, and multiple carriers can be used for networking. Before describing the differences between networking modes, it is useful to look at the concept of a traditional cell. In TD-SCDMA, each carrier section is an independent cell by default. The Uu interface between user equipment and radio access network is configured and operated only in one carrier. To establish a lub interface cell, one cell is configured with an absolute carrier. For multiple carriers, each frequency is regarded as a logic cell. In the case of three sectors and three carrier frequencies, there are nine logic cells that are independently operated. Each cell transmits pilot frequency and broadcast information. Nine carrier frequencies must be configured with nine complete public channels. Broadcast Channel (BCH), Forward Access Channel (FACH) and Paging Channel (PCH) are omnidirectional channels. Therefore, for multicarrier configuration in traditional cell mode, shared frequency networking and difference-frequency networking [1] are typically used.

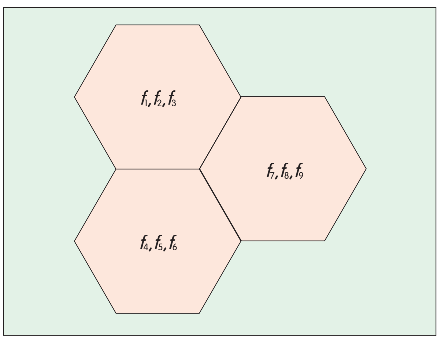
1.1 Shared Frequency Networking Mode

In shared frequency networking, each cell has the same number of carriers and these carriers are the same. Each carrier is an independent logic cell with its own common control channel, downlink pilot channel and independent broadcast channel. On the 10 MHz bandwidth, TD-SCDMA supports up to six carriers. Configuration of shared frequency networking is shown in Fig. 1.

Using the shared frequency networking mode, band utilization can be improved. On the 15 MHz bandwidth, nine carriers are supported, and the station model can be S9/9/9. Multiple logic cells can exist in the same physical environment. On the service channel, service quality can be ensured through intelligent smart and joint detection technology. But the broadcast channel uses omnidirectional transmission. Therefore, interference between carrier



▲Figure 1. Carriers distribution of the shared frequency networking.



▲Figure 2. Carriers distribution of the difference frequency networking.

frequencies is serious, and can significantly impact system performance and capacity.

1.2 Difference Frequency Networking Mode

Compared to shared frequency networking mode, difference frequency networking mode is used to distribute different carriers in adjacent cells. In the case of 15 MHz bandwidth, TD-SCDMA contains nine carriers. As shown in Fig. 2, the maximum station model is S3/3/3.

In difference networking mode, users of the same carrier can be separated to increase the reuse distance and to reduce interference between frequencies. In this way, the system performance and capacity are improved.

1.3 Multifrequency Technology

Due to defects in these two networking modes, Multifrequency technology is proposed. If multiple carrier frequencies exist, one can be selected from the assigned Multifrequency carriers to serve as the primary carrier of the whole cell. The remaining carriers are secondary carriers. Cell division in multifrequency technology is different from that in traditional cells: for multifrequency technology, multiple carrier frequencies of the same section belong to the same logic cell. The primary carrier and secondary carrier use the same scramble code and training sequence code. As a result, multiple carriers in the same cell have the same cell ID. The common control channel is configured on the primary carrier, and on the secondary carrier there is no common control channel. The service channel is configured on both the primary carrier and secondary carrier. Multiple time slot service should be configured on the same carrier to reduce complexity in terminal implementation. Uplink and downlink of the same user are configured on the same carrier. Switch points of the primary carrier and secondary carrier are the same. This restriction is caused by characteristics of the transceiver. If the uplink and downlink switch points of the primary carrier and secondary frequency are different, the base station needs to transmit and receive signals at the same time in some time slots. As a result, transmission signals of the base station can be properly received but other signals cannot. Multifrequency technology can therefore improve system performance and raise frequency spectrum utilization. There are two implementation modes for multifrequency technology: Multicarrier joint networking (of shared frequency and difference frequency) and multicarrier shared frequency networking. For each networking mode, there are three frequency plans covering 15 MHz, 10 MHz and 5 MHz [2].

1.3.1 Multicarrier Joint Networking of Shared Frequency and Difference Frequency

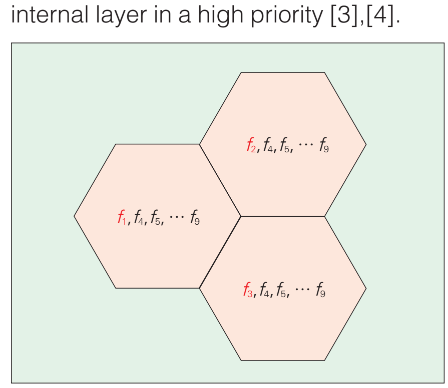
In this networking mode, difference frequency mode is used by major carrier of adjacent cells, and shared frequency mode is used by secondary carriers. For 15 MHz bandwidth, three carriers are selected as the primary carrier, and six are selected as secondary carriers. The maximum station model is S7/7/7. As shown in Fig. 3, the primary carrier is red and the six secondary carriers are black.

1.3.2 Multicarrier Shared Frequency Networking

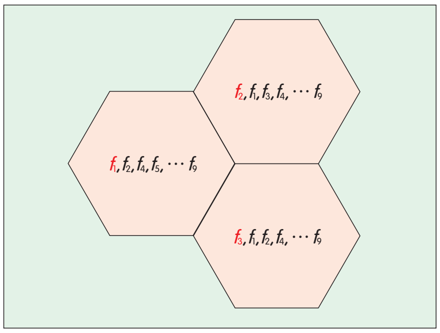
In this networking mode, frequency spectrum can be utilized to the greatest possible extent. Both the primary carriers of adjacent cells and secondary carriers use difference frequency mode. Crossover of adjacent cell main frequency points is contained in the secondary carrier. For 15 MHz bandwidth, the maximum station model is S9/9/9, as shown in Fig. 4.

1.3.3 Carrier Planning

For a network with multiple carriers, carrier planning is very important. Reasonable planning can improve system performance. Hierarchical planning based on multifrequency technology and concentric circle technology of multicarrier shared frequency networking is recommended. This improves system performance and reduces shared frequency interference on the service channel. The cell is divided into two layers, from the center to the margin. Each layer uses a different carrier allocation scheme. The primary carriers of adjacent cells use difference frequency networking modes, and the primary carriers with pilot channel cover the whole cell. Other carriers are called secondary carriers. Compared to the primary carrier, the traffic channel of secondary carriers works for the users of the internal layer in a high priority [3],[4].



▲Figure 3. Joint networking of shared frequency and difference frequency.

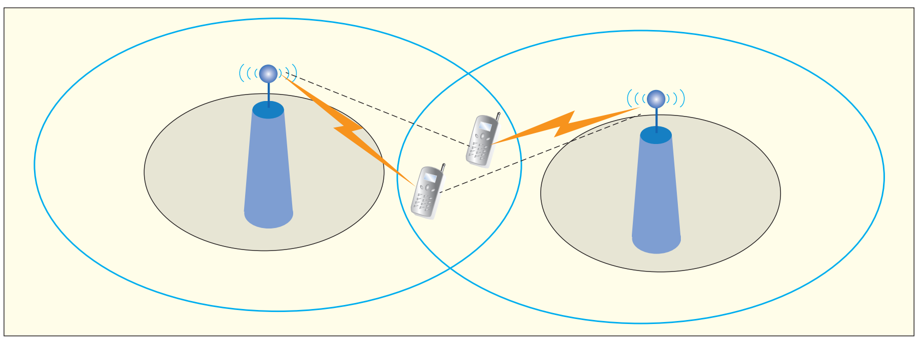


▲Figure 4. Multicarrier shared frequency networking.

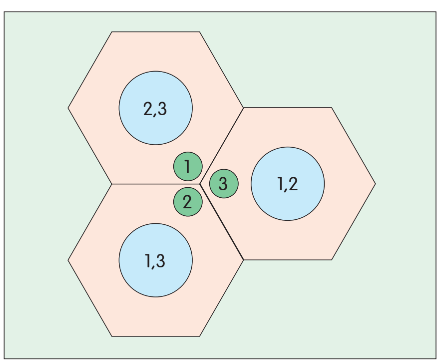
for the 5 MHz frequency bandwidth with three carriers. The typical topology model defined in the Universal Mobile Telecommunication System 30.03 (UMTS 30.03) is used and each cell has three carriers. The primary carriers of adjacent cells are different and each primary carrier of one cell is contained in the secondary carriers of another adjacent cell [5].

As shown in Fig. 6, the red rings divide the cell into two layers. The secondary carriers of each cell are contracted within the internal layer to provide resources for internal layer users. External layer users access the primary carrier point first. Adjacent cells use difference frequency switching, and shared frequency users are isolated.

In this simulation, a mobile model is used (with a speed of 12 km/h. For other parameters, refer to the protocol). The handover algorithm is based on the electrical level, and the handover redundancy value is 1 dB. Open-loop, closed-loop and outer-loop power control algorithm are enabled. Load control is also enabled.



▲Figure 5. Communication between adjacent cells.



▲Figure 6. Cell carrier planning.

As shown in fig. 5, the secondary carrier (internal layer of the cell) is contracted within the yellow area, and the white space is the external area of the cell. When multifrequency technology with secondary carrier contraction is used for hierarchical carrier planning, shared frequency interference of users in the crossover area of adjacent cells is reduced.

2 Simulation Verification

2.1 Simulation

A hierarchical planning model is set

2.2 Major Simulation Parameters

Nine hundred users are placed in the simulation, and each call lasts 200 s. Other parameters are shown in Table 1.

3 Performance Result

3.1 Performance Comparison

When compared with a scheme that does not use hierarchical frequency planning, the performance of the new scheme is as shown in Table 2. When the new scheme is used, transmission power, received interference and call drop rate are reduced. System performance is also significantly improved.

The contraction area of the secondary carrier area is different. When the internal radius is changed, system performance under the new scheme also changes. As the internal radius is contracted, the reuse distance of the shared carrier increases, and interference of the shared frequency decreases. Transmission power and call drop rate are improved. The results of the simulation are listed in Table 3.

3.2 Improvement of New Algorithm

In simulation, hierarchical carrier planning improves system performance. But when users are at the cell edge or in the external layer of the cell, they access the primary carrier first until it cannot provide resources. At this point, the primary carrier becomes overloaded while the secondary carrier is underloaded or not loaded at all. The

▼Table 1. Simulation assumption

	Downlink	Uplink
Transmission Power (Peak Value) (dBm)	Maximum Transmission Power of the Base Station is 30	Maximum Transmission Power is 21
Gain of the Receiving Antenna (dBi)	0	Beam Forming Gain of Smart Antenna is 7 (Omnidirection)
Gain of the Transmitting Antenna (dBi)	Beam Forming Gain of Smart Antenna is 7 (Omnidirection)	0
The Signal-to-Linterference Ratio (dB)	-2.5 (Voice Service)	-2.5
Thermal Noise Power (dBm)	-104	-106
Lognormal Fading (dB)	Base Station-User Equipment: 10	
Power Control Mode	Based on the Signal-to-Interference Ratio	Based on the Signal-to-Interference Ratio
Multiple-User Factor	0	0.78
Non-Orthogonal Factor	0.2	0

▼Table 2. Performance result comparison of new and old scheme

Test Item	Old Scheme	New Scheme	Test Item	Old Scheme	New Scheme
Call Drop Rate ↓	7.5%	0.22%	Successful Access Rate	100%	100%
Base Station Downlink Transmission Power ↓ (dBm)	16.7	7.25	User Equipment Uplink Transmission Power ↓ (dBm)	11.4	−0.94
Downlink ISCP ↓ (dBm)	−74.8	−88.31	Uplink ISCP ↓ (dBm)	−79.8	−93.9
Downlink BLER ↓	36.71%	1.56%	Uplink BLER ↓	23.65%	3.46%

ISCP: Interferencing Signal Code Power BLER: Block Error Rate

▼Table 3. Performance analysis in different internal area radius

Test Item	333	366	400	500
Call Drop Rate ↓ (%)	0.22	0.44	1.5	7.5
Base Station Downlink Transmission Power ↓ (dBm)	7.25	9.8	10.7	16.65
Downlink ISCP ↓ (dBm)	−88.31	−84.60	−81.38	−74.8
User Equipment Uplink Transmission Power ↓ (dBm)	−0.94	1.49	5.59	11.37
Uplink ISCP ↓ (dBm)	−93.89	−91.61	−85.49	−79.75
Uplink BLER ↓ (%)	3.46	4.46	9.91	23.65
Downlink BLER ↓ (%)	1.56	4.02	11.26	36.71

ISCP: Interferencing Signal Code Power BLER: Block Error Rate

▼Table 4. Comparison after the new scheme is improved

Test Item	Improved New Scheme	New Scheme	Test Item	Improved New Scheme	New Scheme
Base Station Downlink Transmission Power ↓ (dBm)	−1.58	−1.2295	User Equipment Uplink Transmission Power ↓ (dBm)	−3.34	−3.06
Downlink ISCP ↓ (dBm)	−101.2	−100.931	Uplink ISCP ↓ (dBm)	−103.0	−102.7

ISCP: Interferencing Signal Code Power

advantages of a new algorithm cannot be demonstrated and calls will be dropped. An improved scheme[6] that can solve the special distribution scenario needs to be found. In the preceding analysis, the algorithm is improved. One carrier is selected from the secondary carrier to serve as the middle one. When primary carrier resources are sufficient, the middle carrier serves as the normal secondary carrier and is contained within the internal cell. When the primary carrier resources are limited, users who excessively access resources may affect system performance. The middle carrier can supplement the primary carrier to provide resources to external layer users. As a result, shared frequency interference can be isolated and system performance in a special user distribution scenario can be ensured.

If the simulation involves two opposite cells each with a radius of 50 meters, the maximum transmission power of R4 time slot is 30 dBm. Each cell has three carriers, and 40 users are placed in the external layer of the cell. The simulation results are listed in Table 4.

The middle carrier of adjacent cells should adhere to the principle that shared frequency interference should be isolated. According to the simulation results, reasonable selection of the middle carrier can improve system performance and save power.

4 Conclusion

From the above simulations, some conclusions can be drawn. The secondary carrier area should be contracted in order to perform networking mode simulation of hierarchical carrier planning. After the shared frequency interference is isolated, system performance is significantly improved. This is useful for the study of multifrequency networking mode and actual networking scheme design. Different user distribution and environment will affect the networking scheme. In the future, a suitable scheme will need to be worked out according to theoretical analysis and actual conditions.

References:

[1] Muan Duan, "TD-SCDMA Frequency Planning and Networking Models," *Telecommunications Technology*, No.11, pp.98-101,Nov.2006.
[2] Hui Xie, Chunjiang Zhu, Rupeng Xu, "TD-SCDMA Wireless Network Planning," *Telecom Engineering Technics and Standardization*, No.10, pp.53-56, Oct.2006.
[3] Zhuo Wang, Lijun Song, Shunyong Yu, "TD-SCDMA System Frequency Planning Analysis," *Telecommunications Technology*, No.5, pp.119-123, May 2006.
[4] Mugen Peng,Wenbo Wang et al., TD-SCDMA Mobile Communication System. Beijing: China Machine Press,2006.
[5] Selection procedures for the choice of radio transmission technologies of the UMTS, UMTS 30.03 v3.20. 1998.
[6] Rong Ren, "Analysis of TD-SCDMA Networking," *Telecommunications Technology*, No.11, pp.29-32, Nov.2005.

Biographies

Min Jin works in the Wireless Signal Processing and Network laboratory at the Beijing University of Posts and Telecommunications (BUPT). He is engaged in researching TD-SCDMA system-level simulation, wireless resource management algorithm, networking planning, and LTE-A physical layer technologies. He has published one article and has applied for one patent.

Wenbo Wang is a professor and doctoral supervisor at BUPT. He is mainly engaged in researching mobile communications radio transmission theory and technology, wireless communication network theory, digital signal processing, and software defined radio theory. He has participated in many state and national science and technology projects, and has received many awards from provincial departments. He receives a special allowance granted by the State Council and is supported by the "New Century Talented Persons Project" of the Ministry of Education. He has published more than 200 articles and 10 books.

Mugen Pen is an associate professor at BUPT. He is engaged in researching signal processing and the key theories, and networking policies of the wireless communication system. His research is now focused on wireless network cooperation information theory, wireless resource management algorithm and protocol design in next-generation wireless broadband, multi-hop wireless relay networking, and network coding technology. He has recently published about 10 books, has translated two books and more than 80 articles, and has applied for more than 20 patents.

ZP−CI/OFDM: A Power Efficient Wireless Transmission Technology

Abstract: Low power efficiency is a deficiency in traditional Orthogonal Frequency Division Multiplexing (OFDM) systems. To counter this problem, a new wireless transmission technology based on Zero−Padding Carrier Interferometry OFDM (ZP−CI/OFDM) is proposed. In a ZP−CI/OFDM system, transmission symbols are spread to all OFDM subcarriers via carrier interferometry codes. This reduces the Peak−to−Average Power Ratio (PAPR) that traditional OFDM suffers and also exploits frequency diversity gain. By zero−padding at the transmitter, advanced receiver technologies can be adopted for ZP−CI/OFDM so that frequency diversity gain can be further utilized and the power efficiency of the system is improved.

Keywords: power efficiency; carrier interferometry; orthogonal frequency division multiplex; zero−padding; frequency diversity gain

Pei Gao
Xiaohu Chen
Jun Wang
(National Key Laboratory of Science and Technology on Communications, the University of Electronic Science and Technology of China)

Orthogonal Frequency Division Multiplexing (OFDM) is based on Multi−Carrier Modulation (MCM). In OFDM, the subcarriers are overlapped but do not affect each other so that the data can be efficiently transmitted based on frequency division multiplexing. This technology has many advantages: Intersymbol Interference (ISI) is countered, spectum can be efficiently utilized, and data can be transmitted at high speed. As a result, it has attracted much attention [1].

In the 1980s, OFDM began to be applied in the telecommunications industry. In the 1990s, it was widely applied in video and audio broadcasting services such as Asymmetric Digital Subscriber Line (ADSL), Very High bit rate Digital Subscriber Line (VHDSL), Digital Audio Broadcasting (DAB), and Digital Video Broadcasting (DVB) [2]. In 1999, IEEE approved the 802.11a standard for 5 GHz wireless local area networks [3], employing OFDM−based physical layer transmission. In the IEEE 802.16 standard [4], OFDM is treated as a basic technology for the physical layer. In a Long Term Evolution (LTE) system, Orthogonal Frequency−Division Multiple Access (OFDMA) is used on the forward link, and Carrier Frequency Division Multiple Access (SC−FDMA) is used on the reverse link [5]. The IEEE 802.15.3a standard for short−distance communication also treats OFDM as an alternative to Ultra−Wideband (UWB) technology [6]. In short, OFDM has become a mainstream transmission technology in broadband wireless communications.

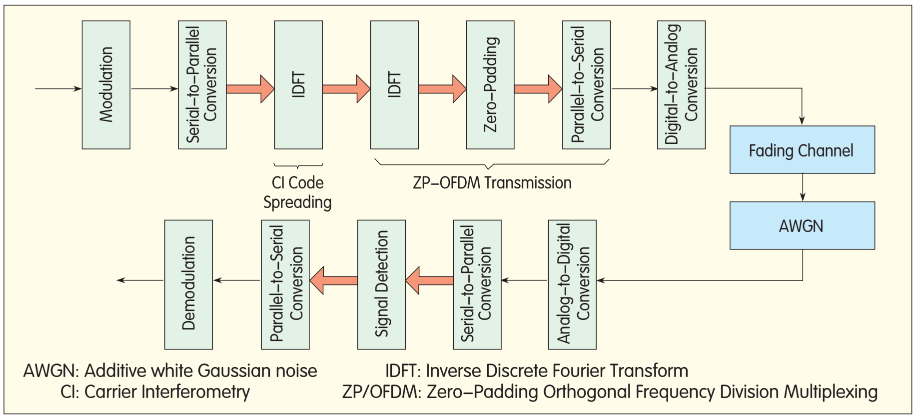
However, OFDM has some defects:

(1) An OFDM system outputs the combined signals of multiple independent subcarriers, and the synthesized signal produces much higher Peak−to−Average Power Ratio (PAPR) compared to signals of single−carrier systems. High PAPR puts high requirements on the linearity of the Power Amplifier (PA) of the transmitter and reduces the transmitter's power efficiency.

(2) By changing the frequency selective fading channel to flat fading sub−channels, an OFDM system can counter ISI and simplify equalization processing at the receiver. However, OFDM is deprived of frequency (multipath) diversity gain when deep fading occurs in some subcarriers because data symbols on these subcarrier are extremely difficult to detect. This limits Bit Error Rate (BER) performance of the OFDM system and reduces power efficiency.

To solve the problem of low power efficiency in traditional OFDM systems, Wiegandt et al. introduced Carrier Interferometry (CI) codes into the OFDM system and proposed an enhanced OFDM transmission technology called CI/OFDM [7],[8]. Unlike traditional OFDM systems, CI/OFDM systems do not transmit low−speed concurrent data via

This work is supported by the National Natural Science Foundation of China under Grant No. 61071102.



▲Figure 1. ZP-CI/OFDM system model.

respective subcarriers but employs orthogonal CI codes to spread the data over all subcarriers for concurrent transmission. Hence, CI/OFDM brings about frequency diversity gain and enhances the system's BER performance without sacrificing transmission rate. Moreover, in the time domain, CI codes enable the peaks of time-domain waveforms to be evenly staggered. This is unlike OFDM systems, where the output comprises the addition of many random sine signals. Thus the problem of high PAPR is eliminated.

Traditional OFDM eliminates ISI by using Cyclic Prefix (CP) as a guard interval. Recent research shows that an OFDM system using Zero-Padding (ZP), that is ZP-OFDM, can recover transmission symbols in the case of deep fading channel. Therefore, ZP-OFDM has better BER performance than traditional CP-based OFDM [9].

1 System Model

Fig. 1 shows the ZP-CI/OFDM system model. At the transmitter, Inverse Discrete Fourier Transform (IDFT) is used for spreading CI code [10]. Data is modulated via an N-point IDFT onto each subcarrier, and N_g zeros are added at the end of the data symbol as a guard interval for realizing ZP-OFDM-based transmission. At the receiver, the signals can be detected in either frequency or time domain, thus frequency diversity gain can be fully used and the system's power efficiency can be improved.

2 Receiver Technologies

In a ZP-CI/OFDM system, signal detection technology at the receiver is critical for power efficiency. Three signal reception models are represented by key signal detection technologies: Frequency-domain Minimum Mean Square Error (MMSE) detection, time-domain MMSE detection, and nonlinear detection.

2.1 Frequency-Domain MMSE Detection

Frequency-Domain MMSE (FDMMSE) detection is designed for the frequency-domain signal reception model. The receiver converts received time domain symbols into frequency domain symbols via an $(N+N_g)$-point Discrete Fourier Transform (DFT). It then estimates a frequency domain channel matrix H of order $(N+N_g)$, which is a diagonal matrix

$$H=\mathrm{diag}(H_0,H_1,\cdots,H_{N+Ng-1}) \quad (1)$$

$$H_0,H_1,\cdots,H_{N+Ng-1} = F_{N+Ng}(h_0,\cdots,h_L,0,\cdots,0)_{(N+Ng)\times 1} \quad (2)$$

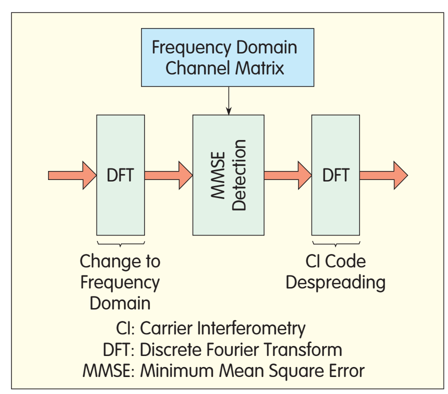
where F_{N+Ng} denotes DFT matrix of order $(N+N_g)$, and $(h_0,\cdots,h_L)$ is the Channel Impulse Response (CIR) vector of fading channel. The receiver uses the matrix H to conduct MMSE detection of frequency domain signals and performs DFT to de-spread CI codes in order to recover the original signals. The schematic diagram of FDMMSE detection is shown in Fig. 2.

2.2 Time-Domain MMSE Detection

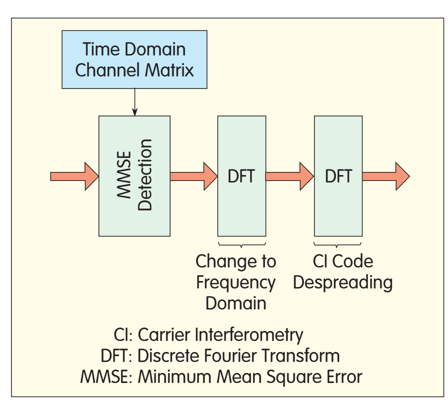
As the FDMMSE detection algorithm cannot make full use of the system's frequency diversity gain, a Time-Domain MMSE (TDMMSE) detection algorithm is further proposed. This technology is based on the time domain signal reception model. By means of time domain channel estimation, the receiver estimates a $(N+Ng)\times N$ time domain channel matrix h, which is a truncated rectangle Toepitz matrix. Then, the matrix h is used to conduct MMSE detection of received time domain signals and performs DFT to convert the signals from time domain to frequency domain. Finally, the receiver performs DFT to de-spread CI codes in order to recover the original signals. The schematic diagram of TDMMSE detection is shown in Fig. 3.

2.3 Nonlinear Detection

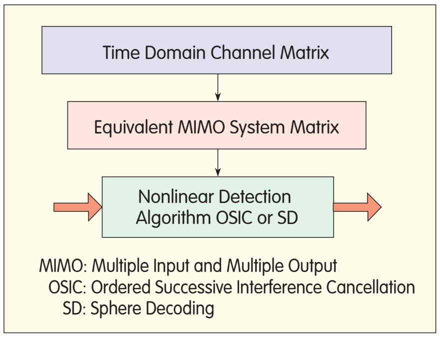
To further make use of diversity gain and improve power efficiency, ZP-CI/OFDM can adopt a nonlinear detection algorithm that is more



▲Figure 2. Implementation of FDMMSE detection.



▲Figure 3. Implementation of TDMMSE detection.



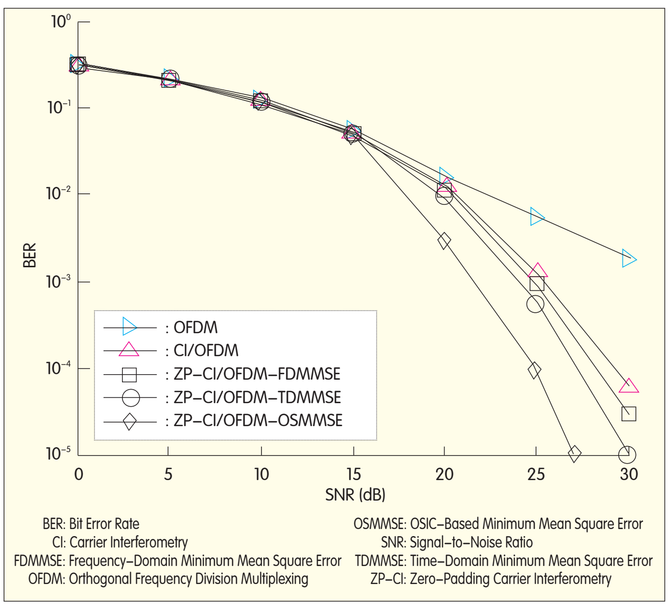
▲Figure 4. Implementation of nonlinear detection.

complicated than time and frequency domain MMSE detection algorithms.

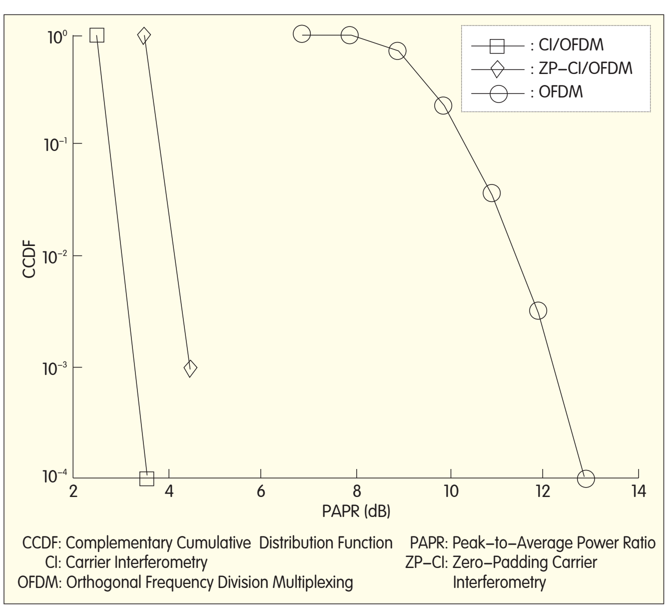
This detection method is based on the prerequisite that the signal reception model can be equivalent to an $N\times(N+N_g)$ Multiple Input and Multiple Output (MIMO) system. On this basis, some nonlinear MIMO detection algorithms can be used for signal detection, thus improving system performance.

The receiver estimates an $N\times(N+N_g)$ time domain channel matrix h. Then it analyzes received signals and generates an equivalent $N\times(N+N_g)$ MIMO matrix Ω, which is from digitally modulated data symbols to received signals,

$$\Omega = hF_N^{-1}F_N^{-1} \quad (3)$$



◀Figure 5. Simulation results of BERs.



◀Figure 6. Simulation results of PAPRs.

where F_N^{-1} is an IDFT matrix of order N. Finally, the receiver uses existing nonlinear detection algorithms such as Ordered Successive Interference Cancellation (OSIC) [11] and Sphere Decoding (SD) [12], for nonlinear detection of the received signals and recovery of the original signals. The schematic diagram of nonlinear detection technology is shown in Fig. 4.

3 Simulation

To verify the high power efficiency of a ZP-CI/OFDM system, simulation tests were conducted to measure the Bit Error Rate (BER) and Peak-to-Average Power Ratio (PAPR) by comparing with traditional CI/OFDM and OFDM.

3.1 BER

Simulation tests were conducted under the following conditions: The channel model was COST207TUx6 [13], the modulation scheme was 16-State Quadrature Amplitude Modulation (16-QAM), the bandwidth was 2.5 MHz, the number of subcarriers was 128, the length of guard interval was 16, and the maximum Doppler shift was 40 Hz.

The simulation results (Fig. 5) demonstrate that BER of ZP-CI/OFDM system is lower than that of CI/OFDM and OFDM systems regardless of which detection technology (FDMMSE, TDMMSE, or OSMMSE) is used. The higher the Signal-to-Noise Ratio (SNR), the more obvious BER gain is. With better use of frequency diversity gain, a ZP-CI/OFDM system uses power more efficiently than an OFDM or CI/OFDM system.

3.2 PAPR

In PAPR simulation tests, Complementary Cumulative Distribution Function (CCDF) is introduced to describe the signal's PAPR. Fig. 6 shows the simulation results of PAPRs of ZP-CI/OFDM, CI/OFDM, and OFDM signals modulated with 16-QAM. CI-OFDM PAPR is the lowest owing to the introduction of CI code. In a ZP-CI/OFDM system, PAPR is slightly higher than that of CI-OFDM because ZP is employed as a guard interval method. However, PAPR is much lower than that of an OFDM system. Therefore, in a ZP-CI/OFDM system, much more BER gain is obtained at the cost of small PAPR.

4 Conclusion

This paper analyzes the advantages and disadvantages of OFDM technology, regarded as a mainstream transmission technology for broadband wireless communications. Low power efficiency is one of the deficiencies that hinder the development of OFDM based wireless transmission technologies. To solve the problem, ZP-CI/OFDM is proposed in this paper. By spreading transmission symbols to all OFDM subcarriers via CI codes, PAPR is reduced, and frequency diversity gain is exploited for ZP-CI/OFDM system. Moreover, by zero-padding at the transmitter, ZP-CI/OFDM can adopt some advanced receiver technologies to make better use of frequency diversity gain and to improve the system's power efficiency.

References

[1] R. Van Nee and R. Prasad, *OFDM for wireless multimedia communications*. London: Artech House, 2000.
[2] H. Rohling, T. May, K. Bruninghaus, and R. Grunheid, "Broadband OFDM radio transmission for multimedia applications," *Proc. IEEE*, vol. 87, no. 10, pp.1778-1789, 1999.
[3] *Draft Supplement to Standard for Telecommunications and Information Exchange between Systems - LAN/MSN Specific Requirements - Part 11: Wireless MAC and PHY Specifications: High Speed Physical Layer in the 5 GHz band*, IEEE P802.11a/D6.0, 1999.
[4] *Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, IEEE 802.16e, 2006.
[5] *Proposed Study Item on Evolved UTRA and UTRAN*, 3GPP RP-040461, 2004.
[6] *TI Physical Layer Proposal: Time-Frequency Interleaved OFDM*, IEEE P802.15-03/141r3-TG3a, 2003.
[7] D. Wiegandt, C. Nassar, W. Zhiqiang. "High-throughput, high-performance OFDM via pseudo-orthogonal carrier interferometry spreading codes", *IEEE Trans. Commun.*, vol. 51, no. 7., pp.1123-1134, 2003.
[8] W. Zhiqiang, W. Zhijin, D. Wiegandt et al. "High performance 64-QAM OFDM via carrier interferometry spreading codes," *Proc. 58th Vehicular Technology Conf. (VTC-Fall'03)*, Orlando, FL, 2003, vol.4, pp. 557-561.
[9] B. Muquet, Z. Wang, G. Giannakis et al. "Cyclic prefixing or zero padding for wireless multicarrier transmissions," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2136-2148, 2002.
[10] F. Xu, R. Xu, and Haixin Sun. "Implementation of carrier interferometry OFDM by using pulse shaping technique in frequency domain," *Proc. 2007 IEEE Int. Workshop on Anti-counterfeiting, Security, and Identification*, Xiamen, China, pp. 319-323.
[11] M. Debbah, B. Muquet, M. De Courville et al. "A MMSE successive inteference cancellation scheme for a new adjustable hybrid spread OFDM system," *Proc. 51st Vehicular Technology Conf. (VTC-Spring'00)*, Tokyo, 2000, vol. 2, pp.745-749.
[12] E. Agrelle, T. Eriksson, A. VARDY et al. "Closet point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, 2002, pp.2201-2214.
[13] "COST 207: Digital Land Mobile Radio Communications," Office for Official Publications of the European Communities, Luxembourg, 1989.

Biographies

Pei Gao (gaopei@uestc.edu.cn) is working towards an M.S. degree at the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. His research interest is signal processing technologies for wireless communications. He has participated in three funded projects, published two papers indexed by EI, and has applied for one Chinese patent.

Xiaohu Chen (cxh4389@163.com) is working towards an M.S. degree at the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. His research interests include signal processing technologies and waveform design for wireless communications. He has participated in three funded projects and applied for one Chinese patent.

Jun Wang (junwang@uestc.edu.cn) is an associate professor at the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. His research interest is signal processing technologies for wireless communications. He has participated in eight funded projects, published 60 papers indexed by SCI/EI, and applied for eight Chinese patents.

Roundup

ZTE Launches zVOA to Empower Converged Voice and Multimedia Services

ZTE Corporation launched its zVOA solution on February 21. This solution enables operators to deploy a unified control network for voice, video and multimedia services over any existing mobile and fixed access network infrastructure.

With the introduction of LTE and WiMAX technologies, widespread deployment of broadband access technologies, and the increasing popularity of multimode smartphones, consumers have come to expect a continuous and consistent service experience in spite of different broadband access points. VoIP and multimedia services are fast becoming a new area of growth for network operators. However achieving cost-efficiency in their networks remains a challenge for many.

ZTE's zVOA solution addresses these issues and utilizes existing traditional voice networks, broadband pipes and end-user resources to offer more multimedia services. The solution is low cost, easy to maintain and easily upgraded. This allows network operators to smoothly deploy VoLTE, VOBB and multimedia networks, and maintain their competitive edge by offering more services to end users.

Key to the zVOA solution is the introduction of the iCX and iMG products. The iCX product integrates standard mobile and fixed softswitches, and IMS call control elements, while the iMG product incorporates media functions with integrated voice, video and other content types.

The solution employs state-of-the-art virtualization technology and Cloud Computing technology on ZTE's advanced ETCA hardware platform and software middleware.

The solution fully complies with a variety of standard interfaces of circuit switched and IMS elements, and can integrate all standard softswitches and IMS functions onto one shelf or even on one single board. (ZTE Corporation)

Research on the Next Generation Naming System

Fuhong Lin
Changjia Chen
(School of Electronic and Information Engineering, Beijing Jiaotong University)

Abstract: Academia has recently proposed new naming systems based on flat Distributed Hash Table (DHT). These naming systems are designed to overcome defects—such as lack of support for data migration and replication—in the Domain Name System (DNS). DHT naming systems have long resolution delay and are not suitable for practical application. This paper introduces two new naming systems that have the advantages of both DNS and DHT systems. The first is a three-layer system based on one-hop DHT and is suitable for small-scale application. The second adopts a hybrid DHT structure, can be implemented in different domains, and can be applied globally. Theoretical analyses demonstrate that these two systems are feasible for practical use.

Keywords: naming system; DNS; DHT; naming; name resolution mapping

Currently, names in the Internet are domain names, and their resolution is provided by the Domain Name System (DNS). The goal of DNS is to map an IP address (which is often difficult to remember) into a meaningful character string that is easy to remember [1]. The advantages of DNS have been fully exploited, and DNS has quickly become the largest domain name system in the world. However, the simple design of early DNS means that it is challenged by today's applications.

First, the root DNS servers have too much authority. All domain name resolution requests that cannot be processed by the local DNS server are forwarded to the root servers. These root servers are controlled by the U.S. government. Therefore, if the U.S. government decided to modify the ROOTZONE files in the DNS root servers, some countries would disappear from the Internet [2].

Second, the domain name is meaningful, and legal wrangling over domain ownership is common [3].

Third, the DNS is host-oriented. After data in the host migrates or is replicated, the DNS cannot provide services for the data again [4].

Fourth, loads are unbalanced among DNS servers. The servers responsible for resolving .com domain names have heavier loads than those responsible for resolving .org domain names. Moreover, there is no coordination mechanism for sharing network resources.

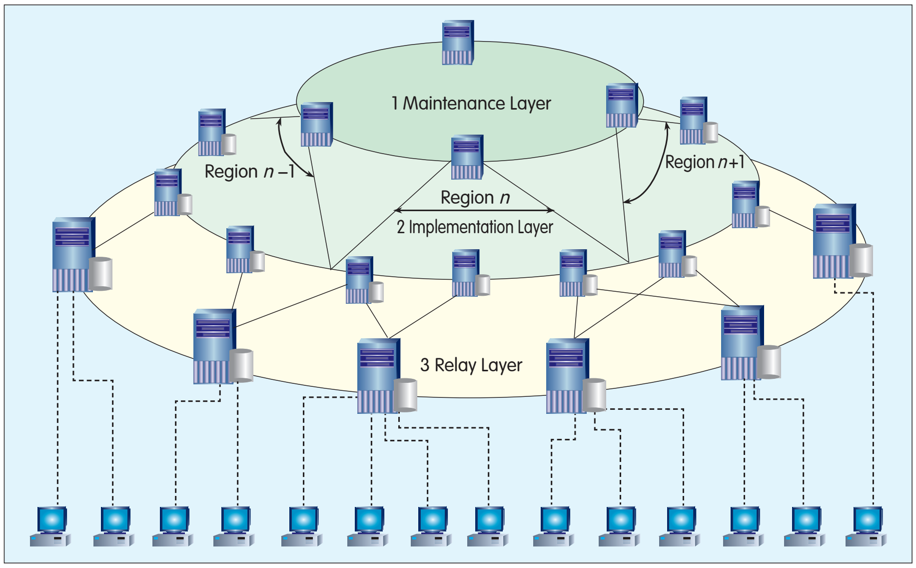
1 New Naming Systems Proposed by Academia

To overcome DNS shortcomings, academia has proposed some new naming systems, of which the naming system introduced in [5] is representative.

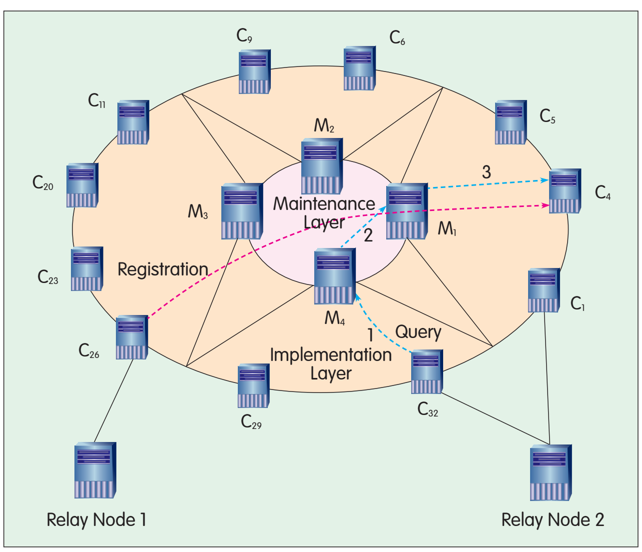
Names are generated in the system using hashing algorithm; meta-data of a resource is first extracted, then a character string is generated by hashing the meta-data. This character string is the resource's network name. Because the name is based on meta-data of the resource, it does not change when the resource migrates or is replicated. It is transparent to all users and can be used to locate the resource at any time. Therefore, the problem is solved whereby services become unavailable once a resource migrates or is replicated.

For name resolution, the system adopts a Distributed Hash Table (DHT) structure. DHT enables a hashed keyword to be mapped to a unique node of the system. In [5], the node in the DHT is regarded as a resolution mapping server. During name registration, meta-data is treated as a keyword and is used to generate a name through hashing. This name is then registered with the DHT. In name resolution, the name is resolved in the DHT and its connection information is returned. Because DHT architecture is flat, if the DHT naming system experiences Denial-of-Service (DoS) attacks, only the attacked server is paralyzed. Other servers are not affected. Such a system is more robust than DNS in the case of DoS attacks. But in a DHT system, there is a maximum delay of log N (take Chord for example), where N is the number of nodes in the system, to resolve a name.

This work was funded by the National Basic Research Program of China ("973" Program) under Grant No. 2007CB307100.



▲Figure 1. One-hop DHT-based naming system.



◀Figure 2. Resource registration and resolution flows.

For a global system, N may be a mega number, so the delay is quite long and the system is not applicable. To solve the long resolution delay problem, two new naming systems are proposed: One-hop DHT-based naming system and overlay naming system. Both have the advantages of the DHT-based naming system introduced in [5].

2 One-Hop DHT-Based Naming System

The one-hop DHT-based naming system is an overlapped structure consisting of three layers: Relay layer, implementation layer and maintenance layer. Fig. 1 shows the structure of one-hop DHT-based naming system. The relay layer mainly controls user access and processes user requests for registration or querying of resource names. The implementation and maintenance layers work together to register or query resource names. The implementation layer maintains a Chord system, while the maintenance layer manages a vector space where a server knows all information of other servers in the layer.

2.1 Implementation Scheme

In the one-hop DHT-based naming system, each node in the relay layer connects to between one and three nodes of the implementation layer—through which the system can be entered for registration or obtaining necessary connection information for services. The implementation layer maintains a Chord ring, and each node maintains two kinds of tables: a parent node table and a routing table. Implementation of the routing table in this system is the same as that in the Chord algorithm [6]. The parent node table contains the ID of the maintenance layer node that manages it. The maintenance layer also maintains a region table and a management table. The region table is used to maintain continuous regions in the implementation layer, and the management table is used to indicate which nodes in the region survive.

An example is taken to briefly explain the entire workflow of the naming system. Fig. 2 illustrates a one-hop DHT-based naming system and the connections between its nodes. The maintenance layer has 4 nodes, the implementation layer has 11 Chord nodes, and there are two relay nodes. Fig. 3 shows three typical tables: Region and management tables of maintenance node M_1, and the parent node table of Chord node C_1.

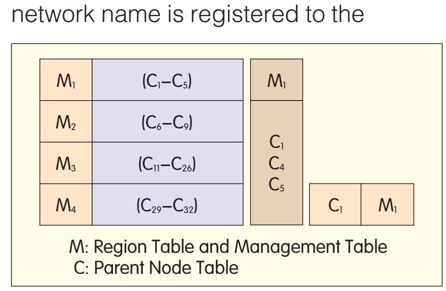
2.1.1 Resource Registration

The resource registration procedure is as follows:

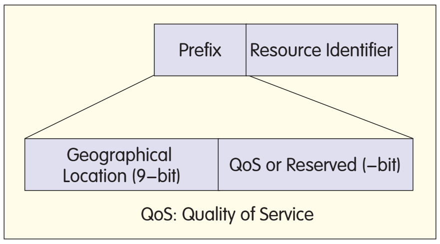
(1) A user sends a resource registration request to a relay node (relay node 1).

(2) The relay node extracts the meta-data of the service and generates a network name (SID1=3) by hashing the meta-data.

(3) Using Chord algorithm, the network name is registered to the



▲Figure 3. Region table, management table and parent node table.



▲Figure 4. Network name structure in overlay naming system.

Chord ring via one access point of the implementation layer (C_{26}). That is to say, information applied to the network name during connection setup is written in another node (C_4).

2.1.2 Resource Resolution

The resource resolution procedure is as follows:

(1) A network name (SID1=3) is generated by hashing the meta–data of a service.

(2) The user sends a resource resolution request to a relay node (relay node 2).

(3) The relay node sends the network name to the upper layer via one access node of the implementation layer (C_{32}) which it connects to.

(4) Upon receiving the service request, C_{32} forwards the request to its parent node M_4 without any processing.

(5) Maintenance node M_4 queries its region table. When it finds the network name 3 is in the region of M_1, it sends the request to M_1.

(6) M_1 queries its management table. When it finds the information of the network name 3 (in C_4), it sets up a connection with C_4 to obtain the connection information necessary for the service and returns it to the user.

(7) The user sets up a connection with the service provider and obtains the service.

The maintenance layer is only involved in the query process because this process is one of the main bottlenecks of network applications. With the maintenance node only processing query requests, query efficiency is improved.

Normally, nodes in the maintenance layer are configured in hot backup and are robust. If a fault occurs, the system no longer sends query messages to the nodes in the maintenance layer but uses Chord algorithm to set up a route in the implementation layer. Nodes in the maintenance layer can be fixed while the system works normally. As a result, the robustness of the entire system is improved.

In the one–hop DHT–based naming system, the maximum number of hops for name resolution is 3, and the query delay is relatively short. But this system is only suitable for small–scale application, and processing capability of servers in the maintenance layer is limited. For large systems, an overlay naming system is necessary.

3 Overlay Naming System

3.1 Name Structure

In an overlay naming system, the network name has a prefix and resource identifier, as shown in Fig. 4. The prefix contains 16 bits that are divided into two parts: the first 9 bits are used to identify the geographical location, and the last 7 bits are used to carry QoS information or are reserved for future expansion. Because there are only around 200 countries and areas in the world, 9 bits are enough to identify them all. The resource identifier is a 160–bit character string obtained from resource meta–data using a hashing function. It uniquely identifies the resource.

The resource identifier is based on resource meta–data and does not change as the resource migrates or is replicated. This characteristic is similar to the naming architecture proposed in [4]; hence, users can always query the resource by name. With geographical location included in the name structure, the resolution range of the resource is shrunk. In addition, the QoS part enables users to select resources that meet their specific needs.

3.2 Name Resolution Mapping Subsystem

The name resolution mapping subsystem consists of two layers: Top layer and bottom layer, as shown in Fig. 5. The top layer is global and adopts DHT architecture for maintaining and managing the names worldwide. The bottom layer is used for a country or area. It adopts DHT architecture for resolving local names and interacting with the top layer and other domains. In the bottom layer, there are ordinary servers and gateway servers. For ordinary servers, name registration and resolution methods are consistent with DHT algorithm; for gateway servers, two new routing tables are added. One is used for routing to top layer servers, and the other is used for routing to gateway servers of other domains.

3.2.1 Resource Registration

Resource registration involves

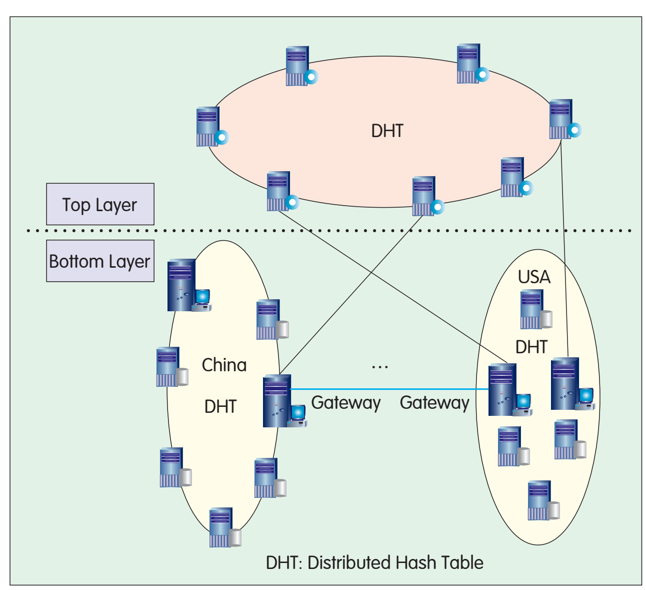


Figure 5.▶ Name resolution mapping of overlay naming system.

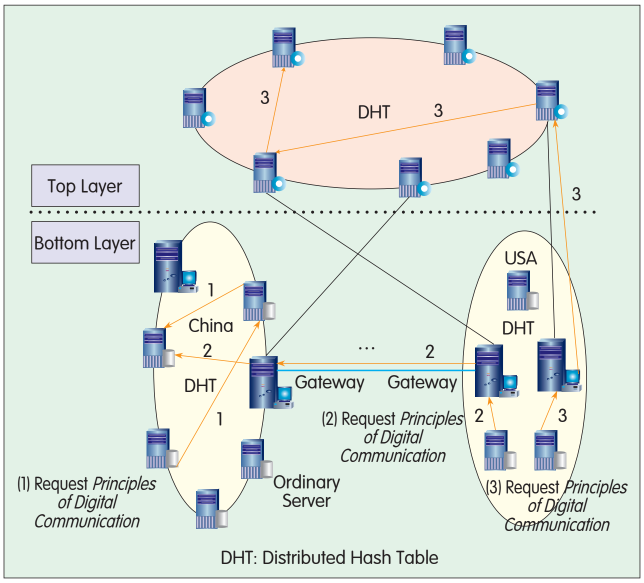


Figure 6
Three name resolution procedures.

generating a name for a resource and registering it with the name resolution subsystem. This procedure is performed by an ordinary server in the bottom layer in six steps:

(1) meta-data of the resource is extracted.

(2) Extracted meta-data is hashed to generate the resource identifier of the resource name.

(3) The country code is entered as the geographical location.

(4) Regional information or QoS information is entered. If the QoS or reserved part is to be reserved, this step is ignored and the part left blank. A legal name is generated for the resource.

(5) Algorithms of the local domain are used to register the name and its connection information to another ordinary server in the domain.

(6) The name and its connection information are registered to the top layer DHT via the nearest gateway server.

Suppose a Chinese service provider wants to register the book *Principles of Digital Communication*. First, meta-data such as principles of digital communication is extracted. Second, a 160-bit character string 101010101111…001 is generated by hashing the meta-data. Third, the country code of China (111000100) is extracted. Fourth, QoS information is entered. For example, QoS class is 3 to 0000111. Finally, all the information above is combined to generate a name: 1110001000000111101010…001.

3.2.2 Resource Resolution

Resource resolution involves obtaining the connection information of a resource from its network name. The resource requester may know the storage location of the resource or not. If the storage location is known, the location may be in the requester's own country or in a foreign country. Fig. 6 illustrates the resolution procedures for these three cases.

Before resolving a resource, a legal name for the requested resource needs to be generated. meta-data of the resource is hashed to generate the resource identifier; the country code is entered if the requester knows the storage location (or 000000000 if the location is not known); and QoS information is entered as the QoS or reserved part.

- If the requested resource is in a domestic server, the resource identifier is used to search the local DHT and find connection information of the requested resource. This occurs at the access point of an ordinary server.
- If the requested resource is in a foreign server, the resource name is sent to the nearest gateway server. The gateway server resolves the geographical location in the name and forwards the request to a gateway server of the country where the resource is located. Upon receiving the request, the gateway server in the foreign country finds the connection information of the resource using DHT routing in its domain.
- If the requester does not know the storage location, the resource name is sent to the nearest gateway server, which directly forwards the request to the top layer DHT. The top layer DHT uses its own routing algorithm to resolve the resource identifier and finds the connection information.

3.3 Delay Analysis

According to [7], network delay is mostly related to the number of hops. Therefore, resolution delay of the overlay naming system can be measured by hops. There are three cases of resource resolution, and the resolution delays for these cases are analyzed in this section.

- For intra-domain resolution, the resolution request is processed in the local domain, and the required maximum number of hops is $\log N_1$ (supposing the servers in the domain are organized as a Chord ring) and N_1 is the number of servers in the domain.
- For inter-domain resolution, the resolution request is first sent to the gateway node of the local domain, which in turn forwards the request to the gateway node of the domain where the resource is stored. The gateway node of the remote domain resolves the name. In this case, the required maximum number of hops is $2+\log N_2$, where N_2 is the number of servers in the remote domain.
- For routing to top domain for resolution, the resolution request is first sent to the gateway node of the local domain. The gateway node then uses the top domain's routing table to resolve the name. In this case, the required maximum number of hops is $\log N_3$, where N_3 is the number of servers in the top domain.

Assuming the percentages of the three cases above are p_1, p_2 and p_3 respectively, the required maximum number of hops for resolution of the system is

$$T_{Hop}=p_1\log N_1+p_2(2+\log N_2)+p_3\log N_3 \quad (1)$$

Suppose each country has the same number of names to be processed, each country has 50 gateway servers, and each country contributes 50 top domain servers. The top DHT system comprises 10,000 servers. If the percentages of intra-domain resolution, inter-domain resolution, and routing to top domain for resolution are 98%, 1%, and 1% respectively, the required maximum number of hops is

$$T_{Hop} = 0.98\log 1000 + 0.01(2+\log 10000) + 0.01\log 10000 = 0.98 \times 13.287 + 0.01 \times 13.287 = 13.307 \quad (2)$$

In [5], Chord ring is used to replace the DNS for name resolution mapping. When there is the same number of servers as overlay naming system; that is, 10,000 in the above example, the required maximum number of hops is

$$T'_{Hop} = \log(200 \times 10000) = 20.9316 \quad (3)$$

Comparing equations (1) and (2), the overlay naming system performs better than the system proposed in [5].

4 Advantages of Two New Naming Systems

Compared with other naming systems, the two naming systems above have the following advantages:

First, they prevent the resolution systems being absolutely controlled by one country, and enable all countries to share network resources. At least one node of the resolution system is deployed in each country, and the entire system serves users around the world. Also, the network name registered with the naming system is a character string that has no real meaning. So one country cannot completely control the system, and the political factor is stripped from network applications.

Second, the new naming systems use the values generated from meta-data as network names for registration and query. Any new service can extract its own meta-data to generate a network name for registration and query. In other words, the new naming systems provide a proper interface for accessing new services. Moreover, the new network name is a nonsemantic hash value without any real meaning. So there will be no wrangling over ownership.

Third, in existing network applications, data replication and migration is inevitable. With existing network technologies, data has to be re-registered with the system and a new domain name assigned after it migrates. It takes a long time for a user to get the new name. As a result, a “http 404 error” message is often seen during Internet surfing. Researchers have developed the http redirection function, which can redirect the data if it migrates within one domain. But for inter-domain migration, there is still no good solution. In the proposed naming systems, a network name is generated from meta-data. When data migrates, the same network name is used to re-register with the system. For users, this process is completely transparent, and migration or replication of data is unnoticeable. Connection information of data can be obtained by resolving the old name.

5 Conclusions

This paper analyzes the shortcomings of DNS, which include meaningful names, incapability of supporting data migration or replication, and unbalanced load. It discusses one DHT-based naming system that has been proposed by academia to overcome the DNS shortcomings. However, new naming systems have long resolution delays and are not suitable for application. Two new naming systems are therefore proposed: One-hop DHT-based naming system and overlay naming system.

The one-hop DHT-based naming system adopts a three-layer ring structure to control user access, to register the resource names, and to resolve the names. But because few upper-layer nodes are responsible for the query, this system is only suitable for small-scale application. In the overlay naming system, name registration and resolution are performed in different domains, and so the system can be used globally. At the bottom layer, each country maintains a DHT-based naming system; while at the top layer, there is a global DHT system. Together, these two systems overcome the shortcomings of the DNS and reduce the resolution delay that affects other DHT-based naming systems. Theoretical analyses demonstrate these two systems are feasible for practical use.

References

[1] P. Mockapetris and K.J. Dunlap, “Development of the domain name system,” in *Proc. Conf. Applications, Technologies, Architectures, and Protocols for Comput. Commun. (SIGCOMM’08)*, Seattle, WA, 2008, pp.123-133.
[2] Jiguang Cao, “New management mechanism of Internet DNS,” *Telecom Network Technology*, vol. 10, pp.8-12, Oct. 2005.
[3] M. Walfish, H. Balakrishnan, and S. Shenker, “Untangling the web from DNS,” in *Proc. 1st Conf. Symp. Networked Syst. Design and Implementation (NSDI’04)*, San Fransisco, CA, 2004, pp.225-238.
[4] H. Balakrishnan, K. Lakshminarayanan, and S. Ratnasamy et al., “A layered naming architecture for the Internet,” in *Proc. Conf. Applications, Technologies, Architectures, and Protocols for Comput. Commun. (SIGCOMM’04)*, Portland, OR, 2004, pp.343-352.
[5] V. Ramasubramanian and G. Emin, “The design and implementation of a next generation name service for the Internet,” in *Proc. Conf. Applications, Technologies, Architectures, and Protocols for Comput. Commun. (SIGCOMM’04)*, Portland, OR, 2004, pp.331-342.
[6] I. Stoica, R. Morris, and D. Liben-Nowell et al, “Chord: a scalable peer-to-peer lookup protocol for Internet applications,” *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp.17-32, 2003.
[7] Fuhong Lin, Changjia Chen, Hengkui Wu et al., “Analysis Time Delay of a Next Generation Name Resolution System,” in *Proc. Conf. 2009 WASE International Inf. Engineering*, Tai Yuan,China,2009.

Biographies

Fuhong Lin (dr.lin.bjtu@139.com) is a doctoral student at the School of Electronic and Information Engineering, Beijing Jiaotong University. His research interests include network architecture and P2P networks. He has participated in several projects funded by the “973” Program, National High Technology Development Program of China (“863” Program), and National Natural Science Foundation of China (NSFC). He has also published more than 10 papers.

Changjia Chen (changjiachen@sina.com), received his doctoral degree from the Department of Electrical Engineering, University of Hawaii. He is a professor and doctoral advisor at the School of Electronic and Information Engineering, Beijing Jiaotong University; director of the Teaching and Research Office of Communications Engineering, Beijing Jiaotong University; a fellow of the Chinese Institute of Electronics; and a senior member of IEEE. His research interests include communications theory, information and encoding theories, and network theory. He has participated in projects funded by “973” Program, “863” Program, NSFC, and the Chinese Ministry of Railways. He has published more than 100 papers.

Adaptive Multiantenna Technology

Huahua Xiao, Dengkui Zhu, and Liujun Hu
(ZTE Corporation)

Abstract: Multiantenna technology can be implemented in several modes. These modes have varying characteristics and are used in different scenarios. This paper introduces Beamforming (BF), Cyclic Delay Diversity (CDD), Spatial Diversity (SD), Spatial Multiplexing (SM), and other multiantenna technologies. It also analyzes various technical features and their application scenarios. An adaptive multiantenna switching algorithm is proposed that chooses a suitable mode for sending data according to the scenario or wireless channel conditions. This switching algorithm improves multiantenna technology and enhances the quality of wireless network communications.

Keywords: BF; CDD; SD; SM; adaptive mode switch

Multiantenna technology is wireless communication technology that deploys multiple antennas at transmitters and receivers. In recent years, multiantenna technology has become an important research issue. Using this technology, power, spatial diversity, spatial multiplexing, and array gain can be achieved while interference [1] is suppressed. Thus, a system's coverage can be enlarged, and link stability and transmission rate can be improved without the need to increase cost too much.

Multiantenna technology works in different ways, for example, Beamforming (BF) [2], Cyclic Delay Diversity(CDD)[3], Spatial Diversity(SD) [4]–[6], Spatial Multiplexing(SM)[7] and a combination of these. Different multiantenna technologies are fit for specific applications. Because the channel environment and location of the receiver change, a single multiantenna technology alone cannot maximize a system's performance. So it is necessary to switch between multiantenna modes to adapt to the ever-changing channel environment.

This work was supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China under Grant No. 2009ZX03003.

1 Introduction to Multiantenna Technology Modes

Every mode has its own characteristics:

(1) SD technology

SD uses signal redundancy to achieve diversity. The transmitter gains diversity by sending orthogonal information set at two different timeslots from two different antennas.

(2) SM technology

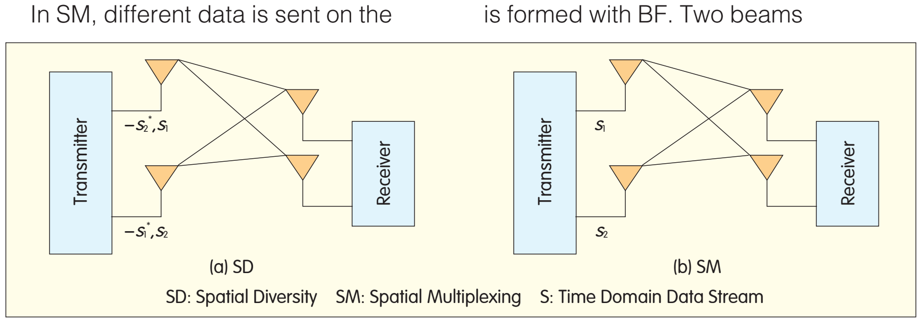
In SM, different data is sent on the same time-frequency resource from different antennas so that spectrum efficiency is multiplied without expending more frequency resource. Fig. 1 illustrates the SM technology. SD and SM are usually referred to as Multi-Input Multi-Output (MIMO) technology.

(3) BF technology

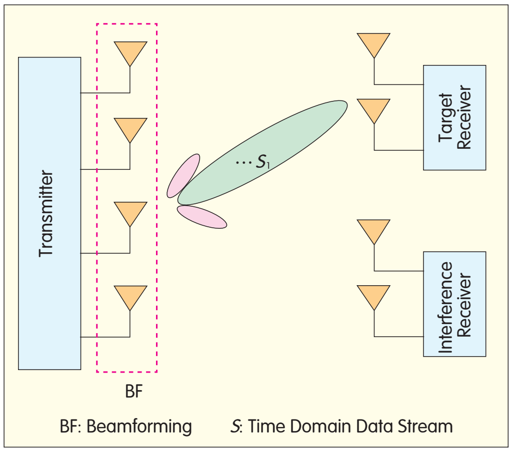
BF is based on an adaptive antenna rationale and uses the antenna array and advanced signal processing algorithm to perform weighted processing on every physical antenna. As shown in Fig. 2, the transmitter performs weighted processing on data stream S_1 and sends it out. As far as the receiver is concerned, the whole antenna array works like a virtual antenna. After weighted processing, the antenna array forms a narrow transmit beam aimed at the target receiver and forms a null point to the direction of the interference reception.

(4) MIMO+BF technology

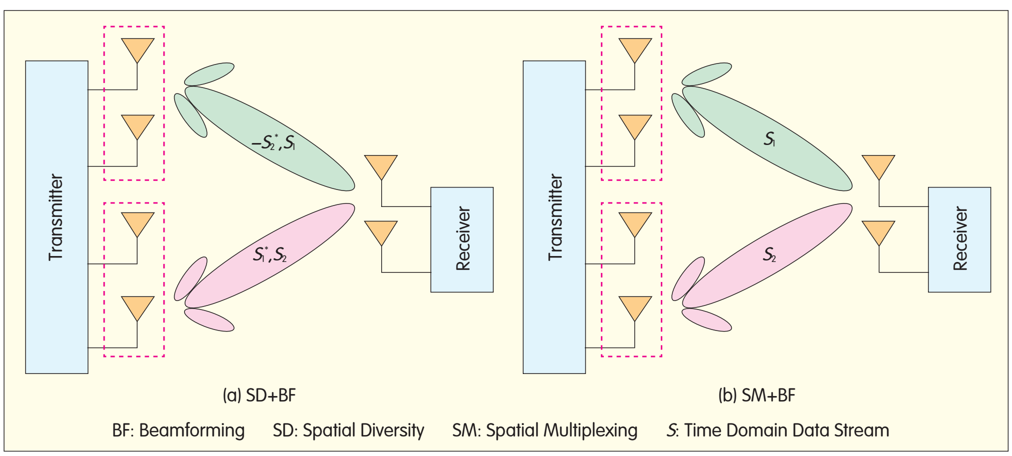
Because only one data stream is transmitted at a time, multiplex is not gained with BF technology. Especially when channel quality is good, transmission rate is not obviously increased with BF. To further improve transmission rate, BF can be combined with MIMO [8], [9]. Combined SD and BF is called SD+BF, and combined SM and BF is called SM+BF. Fig. 3 shows one of the working principles. The four physical antennas at the transmitter are divided into two sub-arrays. On each sub-array, one virtual antenna or beam is formed with BF. Two beams



▲Figure 1.Schematic diagram of SD and SM for two transmission antennas and two reception antennas.



▲Figure 2.Schematic diagram of BF for four transmission antennas and two reception antennas.



▲Figure 3.Schematic diagram of SD+BF and SM+BF for four transmission antennas and two reception antennas.

constitute SD or SM.

(5) CDD technology

CDD is an often-used multiantenna transmit diversity scheme of an Orthogonal Frequency Division Multiplexing (OFDM) system [10]. In CDD, the same frequency domain data is sent on each physical antenna, and different cyclic delays are performed over the OFDM sign of the time domain so that diversity is gained in the frequency domain. Fig. 4 shows the transmitter. The time domain data stream S_1 performs separate cyclic delay δ_i on each physical antenna and then sends them out. δ_i is the amount of cyclic delay, i=1,2,3,4, and δ_1 is usually 0. At the end of CDD processing, the whole antenna array is seen at the receiver as one virtual antenna.

(6) CDD + MIMO technology

Because only one data stream is sent at a time, CDD can be combined with MIMO when the channel condition is good in order to boost transmission rate [11], [12]. SD and CDD can be combined as SD+CDD and SM and CDD can be combined as SM+CDD. Fig. 5 shows one of the working principles. The four physical antennas at the transmitter are divided into two sub-arrays.

2 Comparison of Multiantenna Technologies

(1) Data transmission format

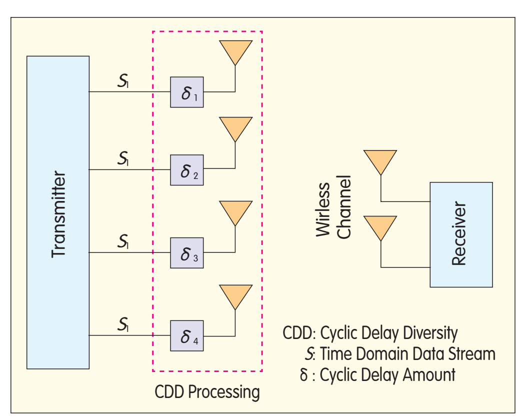
Data sent over physical antennas with diffferent multiantenna technologies is different. Take IEEE 802.16e [4] four antennas for example. Table 1 lists the frequency domain data stream sent by every physical antenna in two neighbor signs and on the same data sub-carrier. SM uses BLAST coding [7], SD uses Alamouti coding, and redundancy is introduced between two Orthogonal Frequency Division Multiple Access (OFDMA) signs. k data sub-carriers on i transmit antennas correspond to BF with weighted value of $w_i(k)$, i=1, 2,3,4. In addition, cyclic delay performed on time domain data is equivalent to the frequency domain data multiplied by a phase rotation

$$D_i(k) = 0.5\times e^{-\frac{j(2\pi k(\delta_i))}{N_F}}[3],$$

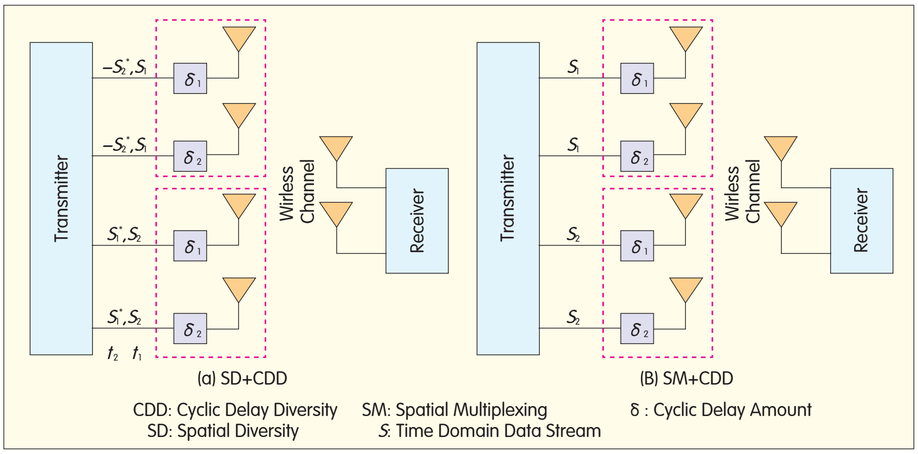
where coefficient 0.5 is the power normalization factor, N_F is the point number of Inverse Fast Fourier Transform (IFFT), k is the sub-carrier index, δ_i is the cyclic delay amount of CDD, and i=1,2,3,4,.The data stream is S_1, S_2, S_3, S...

(2) Characteristics

Usually BF, SD+BF, and SM+BF need to dynamically adjust weighted value depending on the channel status information, which belongs to closed-loop technology. It is also necessary to perform BF on pilot, and therefore a dedicated pilot should be supported. CDD, SD+CDD, and SM+CDD can work when the status of the channel to the transmitter is unknown. This belongs to open-loop technology. SM+BF and SM+CDD can be used to transmit different data streams over different virtual antennas, and if the channel conditions are good, the system´s transmission capability can be enhanced to support high speed data transmission. However, in BF, SD+BF, CDD, and SD+CDD redundancy must be introduced to the spatial dimension in order to gain diversity, which boosts link stability and coverage. In addition, every virtual antenna of SD+BF and SD+CDD can transmit one data stream and introduce redundancy into the time or frequency domains to gain spatial diversity. On average, one data stream is transmitted at a time. BF and CDD send one data stream at a time and apply for applications with higher channel correlation. They are simple to



▲Figure 4. Schematic diagram of CDD for four transmission antennas and two reception antennas.



▲Figure 5. Schematic diagram of combination of CDD and MIMO for four transmission antennas and two reception antennas.

▼Table 1. Data transmission format of different multiantenna technology modes

Multi-Antenna Mode	BF		SD+BF		SM+BF	
	Sign 2	Sign 1	Sign 2	Sign 1	Sign 2	Sign 1
Transmit Antenna 1	$w_1(k)s_2$	$w_1(k)s_1$	$-w_1(k)s_2^*$	$w_1(k)s_1$	$w_1(k)s_3$	$w_1(k)s_1$
Transmit Antenna 2	$w_2(k)s_2$	$w_2(k)s_1$	$-w_2(k)s_2^*$	$w_2(k)s_1$	$w_2(k)s_3$	$w_2(k)s_1$
Transmit Antenna 3	$w_3(k)s_2$	$w_3(k)s_1$	$w_3(k)s_1^*$	$w_3(k)s_2$	$w_3(k)s_4$	$w_3(k)s_2$
Transmit Antenna 4	$w_4(k)s_2$	$w_4(k)s_1$	$w_4(k)s_1^*$	$w_4(k)s_2$	$w_4(k)s_4$	$w_4(k)s_2$
Multi-Antenna Mode	CDD		SD+CDD		SM+CDD	
	Sign 2	Sign 1	Sign 2	Sign 1	Sign 2	Sign 1
Transmit Antenna 1	$D_1(K)S_2$	$D_1(K)S_1$	$-D_1(K)S_2^*$	$D_1(K)S_1$	$D_1(K)S_3$	$D_1(K)S_1$
Transmit Antenna 2	$D_2(K)S_2$	$D_2(K)S_1$	$-D_2(K)S_2^*$	$D_2(K)S_1$	$D_2(K)S_3$	$D_2(K)S_1$
Transmit Antenna 3	$D_3(K)S_2$	$D_3(K)S_1$	$D_3(K)S_1^*$	$D_3(K)S_2$	$D_3(K)S_4$	$D_3(K)S_2$
Transmit Antenna 4	$D_4(K)S_2$	$D_4(K)S_1$	$D_4(K)S_1^*$	$D_4(K)S_2$	$D_4(K)S_4$	$D_4(K)S_2$

BF: Beamforming CDD: Cyclic Delay Diversity SD: Spatial Diversity SM: Spatial Multiplexing

implement, are transparent to the user, and do not need to support MIMO. The most often-used antenna configuration is one in which the transmitter has four or eight antennas and the receiver has one or two antennas. These characteristics are shown in Table 2.

(3) Application scenarios of different multiantenna modes

CDD, SD+CDD, and SM+CDD gain diversity in the frequency domain when channel mulitpath delay is manually introduced. This can be done when the channel status information is unknown. However, in BF, SD+BF, and SM+BF the weighted value of BF needs to be estimated, and the user is required to feedback channel status information or make use of the reciprocity feature of the channel. Therefore, performance is to a great extent affected by the precision and promptness of the estimation of the weighted value. In normal cases where the weighted value is estimated precisely and promptly, performances of BF, SD+BF, and SM+BF are better than those of CDD, SD+CDD, and SM+CDD. If the weighted value is not precise enough because of channel changes or because the user moves too fast, performances may not be as good as those of CDD, SD+CDD, and SM+CDD. In scenarios where channel space correlation is low, SM+BF and SM+CDD can send different data streams over different virtual antennas. However, where channel space correlation is high, BF, SD+BF, CDD, and SD+CDD can gain diversity. Table 3 summarizes the application scenarios of different multiantenna modes [13].

3 Adaptive Mode Switchover

Since every multiantenna mode has its own characteristics and application scenarios, the wireless communication system must adaptively switch between the modes to suit the changing physical location, channel environment, moving speed, and service type of the user. In this way, system performance can be maximized and high-quality communication can occur [14].

In practice, switching between multiantenna modes can present challenges, and there are many factors that can affect the performance of a multiantenna mode [15].

There are many types of mode switchover; switching between BF,

▼Table 2. Characteristics of multiantenna technology modes

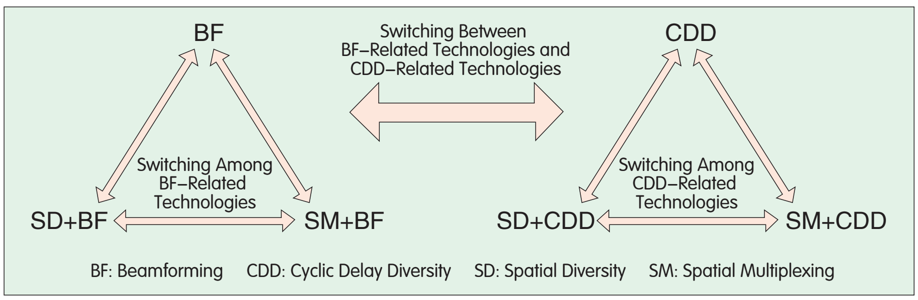
Characteristics	BF	SD+BF	SM+BF	CDD	SD+CDD	SM+CDD
No. of Antennas at Transmitter	2/4/8	4/8	4/8	2/4/8	4/8	4/8
No. of Antennas at Receiver	$\frac{1}{2}$	$\frac{1}{2}$	2	$\frac{1}{2}$	$\frac{1}{2}$	2
Average Qty. of Data Stream	1	1	2	1	1	2
BF Weighted Value	Adjusted Dynamically	Adjusted Dynamically	Adjusted Dynamically	Fixed	Fixed	Fixed
Support of MIMO Needed	No	Yes	Yes	No	Yes	Yes
Support of Dedicated Pilot Needed	Yes	Yes	Yes	No	No	No
Closed/Open Loop	Closed	Closed	Closed	Open	Open	Open
Main Advantage	More Coverage	More Coverage	Higher Traffic	More Coverage	More Coverage	Higher Traffic

BF: Beamforming CDD: Cyclic Delay Diversity MIMO: Multi-Input Multi-Output SD: Spatial Diversity SM: Spatial Multiplexing

▼Table 3. Application scenarios of different multiantenna modes

Application Scenario	BF	SD+BF	SM+BF	CDD	SD+CDD	SM+CDD
High channel Relevancy	Suitable	Comparatively Suitable	Not Suitable	Comparatively Suitable	Comparatively Suitable	Not Suitable
High Moving Speed	Not Suitable	Not Suitable	Not Suitable	Suitable	Suitable	Suitable
Cell Edge	Suitable	Suitable	Not Suitable	Suitable	Suitable	Not Suitable
Cell's Middle Area	Suitable	Suitable	Comparatively Suitable	Suitable	Suitable	Comparatively Suitable
Cell's Inner Area	Not Suitable	Comparatively Suitable	Suitable	Comparatively Suitable	Comparatively Suitable	Suitable

BF: Beamforming CDD: Cyclic Delay Diversity SD: Spatial Diversity SM: Spatial Multiplexing



▲Figure 6. Types of multiantenna mode switchover.

SD+BF, SM+BF, CDD, SD+CDD, and SM+CDD involves at least 15 modes. This makes the switchover algorithm very complex, and further research is required to work out the differences and similarities between difierent switchover types so as to simplify implementation algorithms.

Based on research conducted into the properties of multiantenna technologies in various simulations, three major types of switchover are shown in Fig. 6: BF-related (BF, SD+BF, and SM+BF) switchover, CDD-related (CDD, SD+CDD, and SM+CDD) switchover, and switchover between BF-related technologies and CDD-related technologies.

BF-related or CDD-related technologies are employed depending on the speed of the receiver and the correlation of two neighboring weighted values. For BF-related technology, spectrum efficiency under the SM+BF, SD+BF, and BF modes is calculated, and the data transmission mode with the maximum spectrum efficiency is selected. For CDD-related technology, the spectrum efficiency under SM+CDD, SD+CDD, and CDD is calculated, and the data transmission mode with the maximum spectrum efficiency is selected.

4 Conclusion

This paper introduces the concept of multiantenna modes and analyzes the performance, influential factors, and application scenarios of these modes. It also discusses the algorithm for multiantenna mode switchover. ZTE has not only realized different multiantenna technology modes but has also studied and simulated factors affecting performance. A multiantenna technology mode can be selected depending on the application scenario or channel environment in order to boost system performance to the greatest possible extent.

References

[1] ZTE Corporation, Combination of MIMO and Beamforming Technology for WIMAX, [Online]. Available: http://wwwen.zte.com.cn/en/solutions/wireless/wimax/200912/t20091218_178788.html

[2] M. Kobayashi, G. Caire, D. Gesbert, "Transmit diversity versus opportunistic beamforming in data packet mobile downlink transmission," *IEEE Trans. Commun.*, vol. 55, no. 1, pp.151-157. 2007.

[3] A. Sadeque and M. Saquib, "Delay diversity for correlated MIMO channel," *Proc. 2009 International Waveform Diversity Design Conf. (WD&D'09)*, Orlando, FL, 2009, pp.302-306.

[4] *IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std 802.16eTM rev2-D8, 2008.

[5] L. Zhang and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans.Inf. Theory*, vol. 49, no. 5, pp.1073-1096, 2003.

[6] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp.1451-1458, 1998.

[7] G. Golden, C. Foschini, R. Valenzuela et al., "Detection algorithm and initial laboratory results using V-BLAST space-yime communication architecture," *Electronics Letters*, vol. 35, no. 1, pp.14-16, 1999.

[8] W. Meng, L. Gu, and C. Li, "The Combined beamforming and space-time block coding technique for downlink transmission," *Proc.2005 Int. Conf. Wirelss Networks,Commun.,and Mobile Comput.(WirelessCom'05)*, Maui, HI, vol. 1, pp.481-486.

[9] I. Kim, K. Lee, and J. Chun, "A MIMO antenna structure that combines transmit beamforming and spatial multiplexing," *IEEE. Trans. Wireless Commun.*, vol. 6, no. 3, pp. 775-779, 2007.

[10] X.Tong,T.Luo, OFDM Principle and Applications of Mobile Communication Technology, Beijing: POSTS & TELECOM PRESS. 2000.

[11] C. Yuen, Y. Wu, and S. Sun, "Four transmit diversity schemes for coded OFDM systems with four transmit antennas," *J. Commun.*, vol. 3, no. 4, pp.1-7, 2008.

[12] M. Rahman, S. Das, E. de Carvalho et al., "Spatial multiplexing in OFDM systems with cyclic delay diversity," *Proc. 65th Vehicular Tech. Conf. (VTC-Spring'07)*, Dublin, Ireland, 2007, pp.1491-1495.

[13] Y. Hou,X.Tao. "Architecture of Green Communication-Oriented Base Stations," *ZTE Communications*.vol.16, no.6, PP.16-19, 2010.

[14] W.Zhang, H.Hu,Y.Yang. "Green Wireless Network Technologies," *ZTE Communications*. vol. 16, no. 6, pp. 4-7, 2010.

[15] P.Gao,X.Chen,J.Wang. "ZP-CI/OFDM:A Power Efficient Wireless Transmission System, " *ZTE Communications*. vol. 16, no. 6, pp. 20-23, 2010.

Biographies

Huahua Xiao (xiao.huahua@zte.com.cn) received an M.S. degree in computer software and theories from Sun Yat-Sen University, China. He currently works with ZTE Corporation as a senior engineer in the field of antenna algorithm pre-research. Xiao has applied for more than 30 Chinese and foreign patents in the multiantenna field.

Dengkui Zhu (zhu.dengkui@zte.com.cn) received an M.S. degree in communication and information from Southeast University, China. Zhu currently works with ZTE Corporation as a pre-research manager for the MIMO project in the wireless pre-research department of the R&D system. Zhu has many years of research experience in broadband wireless communications. He has been a member of the pre-research and product implementation team for ZTE's WiMAX and LTE projects. Zhu is responsible for system simulation and key technology research and has a very good understanding of key OFDM and MIMO technologies.

Liujun Hu (hu.liujun@zte.com.cn)graduated from Harbin Engineering University. He now works with ZTE Corporation as a senior engineer and director of the wireless pre-research department. Hu's research interests include mobile communication networks and their key technologies. He has published more than 80 papers and patents.

Green Base Station Solutions and Technology

Zhiping Chen and Licun Wang
(ZTE Corporation)

Abstract: Base station power consumption is the biggest power issue concerning wireless networks. Saving power in base stations is therefore the primary focus in green wireless network development. This paper discusses green base stations in terms of system architecture, base station form, power saving technologies, and green technology applications. It explores effective ways of reducing power consumption in base stations.

Keywords: software defined radio; baseband unit; remote radio unit; high efficiency power amplifier; green power

Environmental protection is a global concern, and for telecom operators and equipment vendors worldwide, developing green, energy-saving technologies for wireless communications is a priority. A base station is an important element of a wireless communications network and often the main focus of power saving in the whole network.

In a wireless communications network, the base station should maintain high-quality coverage. It should also have the potential for upgrade or evolution. As network traffic increases, power consumption increases proportionally to the number of base stations. However, reducing the number of base stations may degrade network quality. So green base stations are proposed. A key issue is how to save energy and reduce power consumption while guaranteeing service and coverage for users and ensuring the base station is capable of evolution.

This paper discusses green base stations in terms of system architecture, base station form, key power-saving technologies, and green technology applications. It aims to find an effective approach to power saving.

1 Multimode Wireless Base Station System

A wireless mobile network is a sophisticated network often with several generations of a system and different frequency bands in coexistence. It contains a 2G system—represented by GSM—and a 3G system—represented by Universal Mobile Telecommunications System (UMTS). Long Term Evolution (LTE) technology has begun to be commercially used. Each generation has its own communications equipment, including a complete set of network elements from base station to core network. For multiple systems to coexist, the traditional network construction approach has been to add new generation devices into the existing network. With more new generation technologies being employed, the number of network devices and auxiliary devices has increased significantly, and this has led to a dramatic increase in network power consumption. Therefore, to reduce power consumption and guarantee QoS in the context of multi-networks, the current network construction approach must be changed.

Based on Software Defined Radio (SDR) technology, the SDR soft base station platform is designed to improve the base station system architecture [1], [2]. The biggest difference between a traditional base station and an SDR soft base station is that the Radio Frequency Unit (RU) of the soft base station is capable of software programming and redefining. So an SDR soft base station can intelligently allocate spectrum and support several standards. By employing broadband multicarrier digital signal processing technology and software configuration, the Radio Frequency (RF) module of the soft base station can support GSM, UMTS, and LTE, as well as transmit and receive multimode RF signals. The baseband module of the soft base station adopts Micro Telecommunications Computing Architecture (MicroTCA) and is characterized by small size, low power consumption, modularization, and good scalability. It can also process baseband signals of several modes, including GSM, UMTS, and LTE.

The SDR soft base station platform enables a telecom operator to combine networks of different modes and different bands into one network. It simplifies network structure and greatly decreases the number of Network Elements (NEs) and auxiliary facilities, thus reducing power consumption base station power consumption.

The 2G/3G swapping project of a

leading telecom operator in Asia-Pacific is a good example of how power consumption can be reduced using the SDR soft base station platform. In the old network, one base station used three cabinets for GSM900, GSM1800, and UMTS2100 devices. Its overall power consumption was 4280 W. After the old base station was swapped with SDR, UMTS900 system was included and power consumption decreased by 57%. In addition, power used by the air conditioner was also reduced because there was less equipment in the room. The SDR base station was effective in saving power.

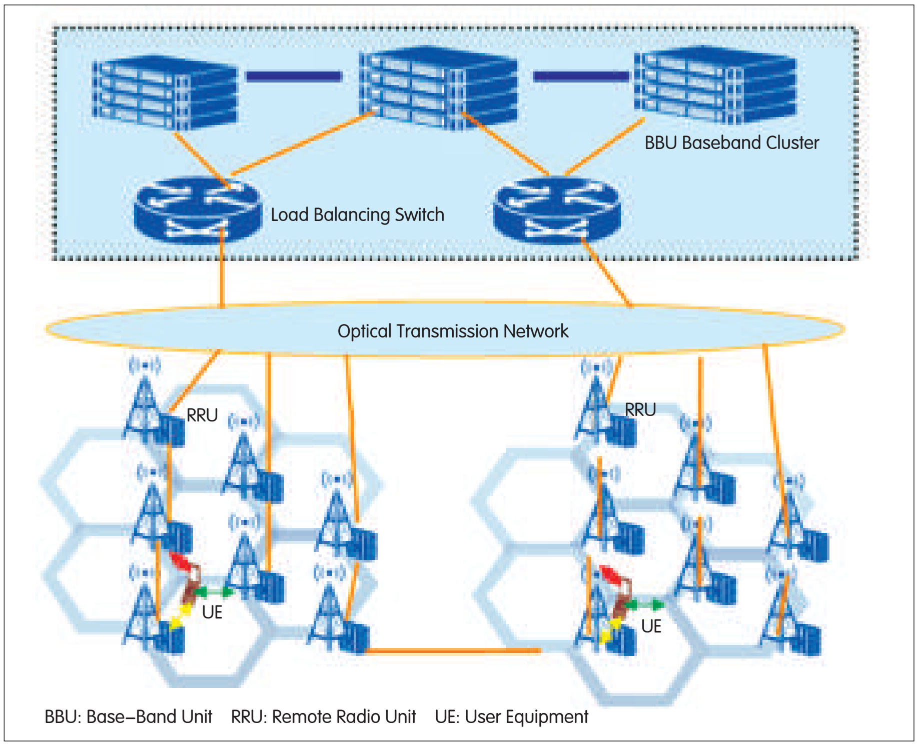
2 Distributed Base Station and Super Baseband Pool

The modular design of an SDR soft base station allows innovation on the base station's form. Two innovative forms are distributed base station and super baseband pool. In distributed base station, the Base Band Unit (BBU) is separated from the Remote Radio Unit (RRU), making network deployment more flexible. A super baseband pool allows baseband resources to be reused and shared and network resources to be used more efficiently.

2.1 Distributed Base Stations

The SDR soft base station platform retains the rack structure form of traditional indoor and outdoor base stations. But it can also launch distributed base stations with BBU and RRU deployed at different places and connected with optical fiber [2],[3].

The RRU can be installed on a roof or in an iron tower and connected to the antenna with a jumper several meters long instead of a feeder cable (which is often tens of meters long). Consequently, the cost of excessive feeder cable is reduced. The Power Amplifier (PA) is also required to output less power, and power of the RRU is saved. With decreased power consumption, the RRU can be ventilated in a natural way, saving on air conditioners or fans. As a result, power consumption and noise of auxiliary facilities is also reduced. The BBU can be inserted into a traditional rack or installed on the wall or on a support to save space. This cuts the cost of land acquisition, equipment room construction, and air conditioning.



▲Figure 1. Wireless access network based on super baseband pool.

2.2 Super Baseband Pool

With the BBU and the RRU separate, multiple BBUs can be placed together and connected to their RRUs with optical fibers. These BBUs form a baseband pool so that baseband resources can be shared, scheduled, and controlled in a centralized way. This solution is called super baseband pooling.

Super baseband pooling changes the form of base station greatly, allowing the baseband's processing capability to be centralized, shared, and virtualized [4]. Processing devices in the baseband pool can be dynamically scheduled to process baseband signals of different RRUs. This enables the base station to adapt to the tidal effect of mobile communications systems and maximize utilization of baseband resources. The RRU can be deployed near the terminal user. This reduces the transmit power on the network and user sides without affecting coverage and reduces power consumption of the whole wireless access network. By centralizing scheduling and control, this solution (Fig. 1) considerably reduces the need for base station equipment rooms, and allows facilities, power supply, and transmission to be shared to the greatest possible extent.

3 Power Saving Technologies for Green Base Stations

Innovation in SDR base station architecture and form brings about greater efficiency in mobile network construction, more reasonable configuration of devices, and maximum sharing of infrastructure and auxiliary devices. Hence, resources are saved. High efficiency PA and intelligent power saving technologies can reduce overall power consumption of the base station even further [5]-[9].

3.1 High Efficient Power Amplifier

Among all components in a base

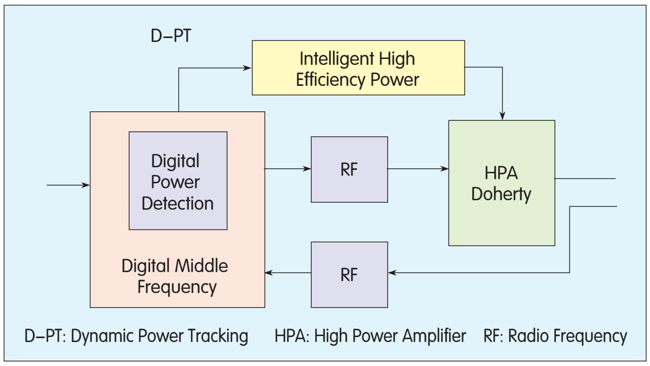


Figure 2. Technical principle of D-PT.

station, the RF unit consumes the most power. And within the RF unit, the PA consumes the most power, accounting for about 80% of the RF's total power consumption. Decreasing power consumption in a base station reduces the heat generated by equipment and the amount of air conditioning needed for cooling. Therefore, improving the PA's efficiency reduces power consumption of the main devices in a base station.

When designing high efficiency PAs, circuit application, component selection, and process innovation must be taken into account. PAs have developed from the expensive feed-forward linear type and class AB high powered type to Doherty Pas, which come with Digital Pre-Distortion (DPD) technology. The PA chip has also developed, from Laterally Diffused Metal Oxide Semiconductor (LDMOS) and GaN to High Voltage Heterojunction Bipolar Transistor (HVHBT). PA efficiency has increased from less than 10% to 45% and is now targeted for over 50%. Currently, amplifiers combining DPD and Doherty dominate in base stations in wireless communications systems.

The demand for continuous improvement in efficiency drives the development of PA technologies. New PA technologies include Envelope Tracking (ET) PA and digital switch PA.

3.2 Intelligent Power Saving Technology

The mobility of mobile users means base station traffic varies dramatically from hour to hour. Intelligent power saving technologies are designed to assess traffic in real-time and take power saving measures based on assessment results. Idle resources can be made dormant by intelligent carrier adjustment or intelligent channel power-off. The PA's power supply can be configured by intelligent PA control or dynamic voltage scaling. Here, dynamic voltage scaling and intelligent carrier adjustment are analyzed.

(1) Dynamic Voltage Scaling

Dynamic voltage scaling is also called Dynamic Power Tracking (D-PT). D-PT involves intelligently controlling the PA's power supply by tracing load changes and using variable voltage. Fig. 2 shows the technical principle of dynamic voltage scaling. When the power output of the PA is high, voltage provided to the PA is large. When the output power falls below a certain threshold, the voltage is lowered too. This ensures the PA works at its best under different power loads and saves power in different configurations. An intelligent and highly efficient power supply system can adjust the power supply as the load changes. When the load is at maximum, power efficiency is as high as 92%. Generally speaking, a base station's power consumption can be reduced by 12% using this technology.

(2) Intelligent Carrier Adjustment

The load of a base station dynamically changes—traffic in peak hours differs greatly from that at other times. In a base station, the number of carriers is usually configured according to peak hour traffic. As a result, in idle hours, the power of some carriers is used in control channels rather than in traffic channels, leading to very low power utilization.

Intelligent carrier adjustment enables a base station to dynamically adjust the number of active carriers based on traffic, and inactive carriers are shut down to reduce power overhead. In an S222 configuration, this technology can reduce base station power consumption by about 40% in idle hours.

4 Application of Green Base Station Technologies

Temperature control devices such as air conditioners are accountable for a considerable amount of power consumption in a base station. Before green base stations are put into use, one of the key issues is how to effectively save power in temperature control system. Solar and wind is now being widely applied in the telecommunication industry to power base stations.

4.1 Intelligent Temperature Control System in the Equipment Room

Fans in wireless devices and air conditioners in equipment rooms consume a lot of energy in order to create a favorable working environment for system devices. An effective way to save energy is to reduce power consumption of fans and air conditioners.

Automatic Control Systems (ACS) can be used for this purpose. Temperature sensors are installed indoors and outdoors to measure temperatures, and then the indoor temperature is controlled with natural wind. Only when the difference between indoor and outdoor temperatures is small and the indoor temperature exceeds a predefined threshold does the ACS run the air conditioner. The structure of the ACS is illustrated in Fig. 3.

ACS can be used independently or with the air conditioner. By taking advantage of natural conditions, the temperature of a base station can be adjusted in all weather. This intelligent temperature control system uses fans instead of air conditioners for ventilation for about 80% of the year. As a result, the annual running time of an air conditioner is greatly reduced. Compared with a traditional equipment

D:\EMAG\2010-05-26/VOL8\COVER.VFT——5PPS/P64

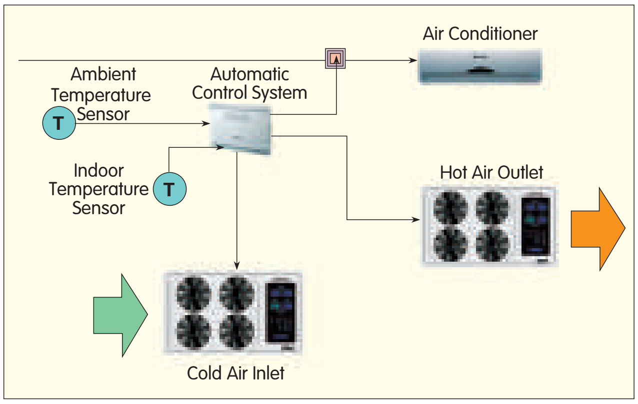


Figure 3. Automatic control system for equipment room temperature.

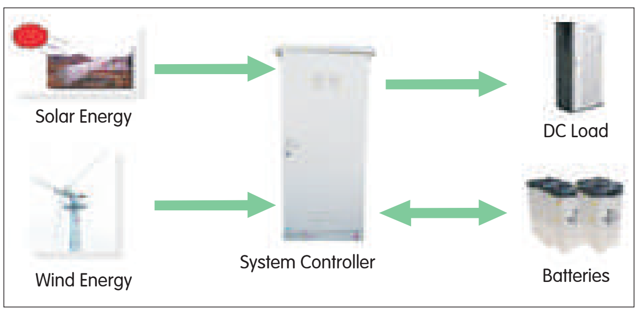


Figure 4. Hybrid power supply solution with wind and solar energy.

room, an ACS-cooled room can save up to 70% energy.

4.2 Power Supply Solution with Green Energy

A sharp decrease in power consumption in a base station makes it possible to replace the traditional electrical power supply with solar or wind energy. Among other solutions, solar and hybrid solar-wind power has gradually been applied in base stations.

Solar and wind generated power is clean, inexhaustible, and cheap. Long-term benefit can be gained from an initial investment, and so it conforms to the global trend towards power saving and emissions reduction. However, these solutions depend on favourable weather conditions. For solar energy, the average daily solar irradiance must reach 4 KWh/m and for wind power, wind speed must be at least 3.5 m/s to make the turbine work.

At present, hybrid wind and solar energy is the most feasible green power solution (Fig. 4). During the day, a solar panel and wind turbine provide power to base station equipment; while at night, equipment is powered by wind and batteries. When there is neither wind nor sunshine, the batteries are used. Wind and solar devices can be flexibly configured based on site situations. In principle, the ratio of wind energy to solar energy can vary from 2:8 to 5:5. But in any case, power supplied using wind cannot exceed 50% of the total power supply.

5 Conclusions

The green base station solution involves base station system architecture, base station form, power saving technologies, and application of green technologies. Using SDR-based architecture and distributed base stations is a different approach to traditional multiband multimode network construction. Using this approach, power consumption in the network can be reduced greatly. Advancement of PA technologies and the introduction of intelligent power saving also improve resource utilization and reduces emissions. Intelligent temperature control and new energy sources make wireless base stations greener. Although reducing power consumption and emissions in a wireless network requires various power saving means and technologies, technical updates and innovations in the base station itself are key for green base stations.

SDR-based BBU+RRU base stations have been deployed globally, and with their environmental protection features, have won the trust of global users. However, innovation in green base stations has not come to an end. Continuous improvement in wireless communications network will bring a better future.

References

[1] Xiyu Wang, Dandan Guo, and Z. Cui, "Soft base station technology in wireless communications Systems," *ZTE Communications*, vol. 8, no. 4, pp. 49-53, 2010

[2] Chunming Jia and Xiaofei Fei, "SDR soft base station expanding 3G network new platform," *World Telecommunications*, no. 12, pp. 71-72, 2008.

[3] Yongjian Mou, "ZTE's distributed soft BTS solution," *Designing Techniques of Posts and Telecommunications*, vol. 29, no. 4, pp. 38-39, 2008.

[4] Qingjie Song and L. Yang, "Research and application of new 3G networking solution based on super baseband goups," *Shangdong Communication Technology*, vol. 30, no. 2, pp. 27-30, 2010.

[5] Yuli Song, Dongxiao Li, and Zhansheng Wu, "Discussion on smart power saving technology of BTS", *Designing Techniques of Posts and Telecommunications*, vol. 31, no. 3, pp. 70-73, 2010.

[6] J. Xiao, J. Li, and Shuangjiang Tan, "Initial analysis on energy saving and emission reduction techniques for base stations," *Designing Techniques of Posts and Telecommunications*, vol. 31, no. 6, pp. 72-76, 2010.

[7] X. Liu and Wenzhou Cai, "Energy-saving base stations constructing ecotype TD network," *Mobile Communications*, Vol. 33, no. 12, pp. 87-89, 2009.

[8] Qianhao Lv, Wenyong Mao, and J. Zhu, "Green GSM network solution leading environmental protection trend," *Designing Techniques of Posts and Telecommunications*, vol. 30, no. 6, pp. 25-30, 2009.

[9] Jinlu Xu, Haibo Yan, and L. Sun, "Solution of saving energy and reducing emissions in mobile communication base station," *Designing Techniques of Posts and Telecommunications*, vol. 30, no. 12, pp. 57-59, 2009.

Biographies

Zhiping Chen (chen.zhiping@zte.com.cn) has a master's degree from Huazhong University of Science and Technology. She is a solution director in the architecture department of Wireless Product Operation Division of ZTE Corporation. She is mainly responsible for technical support and developing integrated solutions for high-end wireless communications markets.

Licun Wang (wang.licun@zte.com.cn) has a Ph.D. from the Institute of Automation of the Chinese Academy of Sciences. He is a solutions architect in the Architecture Department of Wireless Product Operation Division of ZTE Corporation. He is mainly responsible for demand analysis and integrated solution development for high-end wireless communications markets. He has published 30 papers.

The Internet of Things and Ubiquitous Intelligence (1)

Dongliang Xie
Yu Wang

(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications)

Editor's Desk:

The traditional Internet is oriented towards person-to-person connection, whereas the Internet of Things is oriented towards connection of inanimate objects. As such, the Internet of Things covers a larger range of connections and involves more semantics. Internet and telecom networks are focused on information transfer, while the Internet of Things is focused on information services. By combining sensor networks, the Internet, telecom networks, and cloud computing platforms, the Internet of Things can sense, recognize, affect, and control the physical world. The physical world can be unified with the virtual world and human perception. This lecture discusses Internet of Things technology from three perspectives: Ubiquitous information sensing, ubiquitous network convergence, and intelligent information services. In this part, we will introduce the technical characteristics of the Internet of Things and sensor networks, the development background of sensor networks, and key technologies of sensor networks.

1 Introduction

1.1 Technical Characteristics of the Internet of Things and Sensor Networks

Inspired by new technologies in computing, microelectronics, communications networking, and man-machine interaction, society has shifted its focus from network connection services to ubiquitous information services. With this shift, the Internet of Things has developed rapidly. However, as its technical connotations and denotations are continuously evolving, there is currently no uniform, integral, or accurate definition. People are defining the Internet of Things differently during its different stages of development. On November 17, 2005, the ITU released *ITU Internet Report 2005: The Internet of Things* [1], where the Internet of Things was defined as Radio Frequency Identification (RFID)-based thing-to-thing and thing-to-person Internet. But as sensor networks have matured and modern service technologies and new services have emerged, traditional definitions of the Internet of Things no longer satisfy technical and application requirements.

The Internet of Things is generally thought to be a network integrating information collection (using wireless sensor networks), information transmission (using the Internet and telecom, radio, and TV networks), new services and applications based on information service networks are also part of the Internet of Things. Generally speaking, the Internet of Things has typical characteristics:

(1) Ubiquitous information capturing. The sensor network is the engine of the Internet of Things and is at the tip of information sensing and processing. What makes sensor networks distinct from other networks is their capacity to collect massive amounts of diverse information on environment and events in close proximity. This information is obtained from various perspectives and for multiple parameters. The diversity of information, mass quantity of information, and complexity of relationships between information is unprecedented. With node identification, sensor devices can assign specific locations and identifications to a piece of information, which allows the information to be used in future applications.

(2) Reliable information transmission. In the Internet of Things, information categories become much richer and differentiation in Quality of Service (QoS) becomes more complicated than in existing networks. Information service rather than connection service will be a basic operation feature of the Internet of Things. As an infrastructure and support environment for a ubiquitous information society, ubiquitous networks will be an important objective in developing information communication networks. Reliable and effective transmission technologies are required in existing networks for providing ubiquitous and intelligent services and providing people with rich real-world information.

(3) Efficient information applications. The major difference between the Internet of Things and existing networks is ubiquitous intelligence. The richness that comes with masses of sensing information and reliable information

transmission is essential for providing diverse services using the Internet of Things. Processing and utilizing mass sensing information can provide enterprises and the general public with new service modes and experiences—especially ubiquitous services, business operation modes, and management systems that can adapt to context.

This lecture discusses three important components of the Internet of Things: Sensor network collection of information, ubiquitous network-based information transmission, and new service technologies and businesses using collected information. This part focuses on the sensor network.

1.2 Development Background of the Sensor Network

Since the middle of the 20th century, computing technologies and network technologies have together given rise to many new application modes and technical means. But all these fall into the category of traditional person-to-person or person-to-machine interaction. With the rapid development of communication technologies, embedded computing technologies, and microelectronic technologies, the manufacturing costs of electronic products has been steadily decreasing, as predicted by Moore's Law. In the 1990s, the first microintelligent sensor was put into use that had sensing, computing, and communication capabilities and was suitable for different environments. Using the microsensor as the network node—Wireless Sensor Network (WSN) technology—completely changed the way information is acquired [2].

WSN is a potential network. Many simple nodes can be randomly deployed in remote locations and under adverse conditions. In a self-organizing network, all nodes work cooperatively to monitor, sense, and collect information in complicated environments or to monitor objects within the network's distribution area and process the information. In this way, nodes obtain detailed and accurate information and then send it to interested observers. Because it is a bridge between the objective physical world and subjective sensing world, WSN is a new information acquisition and processing technology, a revolution in information sensing and collection.

Now, WSN is a hot research topic in network communications and information processing in the U.S., and efforts to develop it are being stepped up. In its Information and Communication Technologies (ICT) Seventh Framework Programme (FP7) started in 2007, the European Union lists WSN as an important research area in network embedded control systems. In South Korea, the IT839 strategy makes ubiquitous sensor networking one of the three key infrastructures. In Japan, both e-JAPAN and U-JAPAN strategies have sensor networking as a goal for creating a next-generation Information and Communication Technology (ICT) society by 2010.

2 Key Technologies of the Sensor Networks

2.1 Nodes

Nodes are the basis for WSN research and application. All protocols, mechanisms, and algorithms related to WSN must be implemented via nodes. Therefore, node design and implementation determine functions, performance, and investment across the entire network. Most WSNs use general embedded platforms as their nodes; for example, Crossbow's Mica-and Telos-series nodes. These platforms use common CPUs to connect peripheral devices—such as sensors and RF chips—in order to form sensor nodes. Node functions are developed mainly by software. On the one hand, node functions, supported by TinyOS and other operating systems, are relatively easy to develop. On the other, node development is limited because of fixed hardware and node size, and cost. Also, power consumption cannot be reduced through physical design. Nodes have limited processing capability and it is difficult to develop sophisticated programs to counter this.

Most WSN applications rely on critical performance indexes for achieving low power consumption, low cost, small size, easy deployment, and reliability. One practical approach to meeting these demands is to employ System on Chip (SoC) nodes. Highly integrated technology and good physical design can reduce size, cost, and power consumption of nodes, and architecture designed especially for WSN improves its computing efficiency and capability. This is a trend for future node technologies.

2.2 Network Protocol and Networking

In WSN, the protocol stack design determines the self-organization mode and communication performance of network nodes. It also determines interconnection and access methods of heterogeneous networks. Protocol stack design ensures effective energy, scalability and reliable transmission in WSN when there are limited network resources. WSN protocol stack and working technologies fall roughly into two categories: Non IP-based protocol stack and IP-based protocol stack.

Two examples of non IP-based protocol stacks are Zigbee and Sensor-Net. Zigbee is a low consumption, low power, reliable wireless network standard from the Zigbee Alliance [3]. Sensor-Net is an architecture especially designed for the sensor network [4]. It adopts cross-layer design for energy management, system management, time synchronization, discovery, and security. Inter-layer cooperation allows better management and control.

At present, there are only a few IP-based protocol stacks, including NanoStack, PhyNet, and IPv6 micro sensor router. NanoStack is an embedded sensor network software project based on IPv6 over Low power Wireless Personal Area Networks (6LoWPAN) [5]. This protocol stack contains User Datagram Protocol/Internet Control Message Protocol (UDP/ICMP), 6LoWPAN IPv6 protocol, and IEEE 802.15.4 Media Access Control (MAC) protocol.

Because IPv6 allows for much larger packet capacity than the maximum frame capacity specified in IEEE 802.15.4, an adaptation layer is added between the network and data link layers to splice and reorganize IP packets. In this way, low-rate transmission requirements of Wireless Personal Area Networks (WPAN) can be met. PhyNet protocol stack is an industrial solution based on the IETF's 6LoWPAN. It is used for implementing IP-based communication in WPAN. This protocol stack takes the 6LoWPAN standard as its core and supports IEEE 802.15.4 physical layer and MAC layer protocols.

In addition, WSN must support interconnection with other networks for sharing and processing information. Further study into the architecture and self networking technologies of WSN is required.

2.3 Information Processing

Information processing technologies of the sensor network are diverse, ranging from embedded, lightweight information processing technology within a node to distributed network information processing technology between nodes. On the one hand, WSN is cheap, has high observation precision, and can integrate heterogeneous sensor information. These advantages lay the foundation for implementing an ideal information processing technology. On the other hand, with low power consumption, heterogeneous interconnection, ubiquitous cooperation, and limited nodes, WSN has some technical drawbacks for information processing. The progress of research on core information processing technologies for the sensor network is as follows:

- Signal processing, information identification, and information extract technologies within a node: These technologies are used to guarantee the accuracy of signal collection and sensing. Especially for sensor networks with limited nodes, embedded and lightweight signal processing, information identification, and extract technologies are necessary. Node signal processing technology and information identification, information extraction, and application of technology are closely related. These technologies need to tune the node degree of intelligence and lightweight signal processing algorithms constrained by the node resource.
- Distributed cooperative processing technologies between the nodes: The goal of these technologies is to improve a node's intelligence in estimating its surroundings and identifying attributes. Such technologies include cooperative technology among distributed nodes and multisensor information integration technology. With limited resources, a node in the sensor network can only sense local information. Hence, all nodes in the network must cooperate with each other to obtain global information. In a multisensor system, information is extremely diverse, the quantity of information is massive, and relationships between information are complex. Therefore, multisensor information integration technology is the key for information processing in WSN.
- Information compression and coding technologies. These can greatly improve the transmission performance of a multihop link and bring about effective, reliable signal transmission. Packet compression is an effective low consumption communication technology that reduces the length of a packet by information redundancy. It can shorten end-to-end transmission delay of a data packet and reduce conflict in accessing shared channels.

2.4 Security

Due to resource limits, WSN has more security challenges than traditional networks. A low consumption security solution designed for WSN should take into account optimal loads for communication and computing, distribution and update of secure keys, security authentication and anticapture of nodes, and scalability in secure key management. The goal of WSN security technologies is to ensure information security during node communication.

Key management is a basic but important step in security management. To improve security and antiattack performance of the network, a key management mechanism needs to be implemented. Current research into attacks against sensor networks focuses mainly on secure routing technologies. In China, researchers have been studying specific key management protocols—which are still theoretical models—rather than studying sensor network security from a systematic perspective. Moreover, research on secure routing technologies is concentrated on stability and simulation of routing algorithms. Although some security models and objectives have been proposed, there is still no secure routing approach that is feasible for sensor networks.

In China, research into security models for sensor networks has been focused on particular scenarios rather than on a general security solution. Therefore, one area that requires systematic study is general security models for sensor networks, including secure communication protocols, secure key management, secure routing, and access control.

(To be continued)

References:

[1] International Telecommunication Union (2005) ITU Internet Report 2005: The Internet of Things. Geneva: ITU. [Online]. Available: http://www.itu.int/pub/S-POL-IR.IT-2005/e

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, no.8, Aug. 2002.

[3] ZigBee Alliance. [Online]. Available: http://www.zigbee.org/

[4] J. Polastre, J. Hui, P. Levis et al., "A unifying link abstraction for wireless sensor networks," in *Proc. 3rd ACM Conf. Embedded Networked Sensor System (Sensys05)*, San Diego, CA, 2005, pp.

[5] "IPv6 over low power WPAN," datatracker.ietf.org. [Online]. Available: http://datatracker.ietf.org/wg/6lowpan/charter/

Biographies

Dongliang Xie (xiedl@bupt.edu.cn) is a director and associate professor at the Broadband Network Center of State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. He researches wireless and mobile network technologies, wireless sensor networks, mobile Internet QoS, network cooperation, and ubiquitous intelligence. He has published more than 40 papers.

Yu Wang (wang0yu@gmail.com) is a master's student at the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. She researches convergence of wireless sensor network and ubiquitous network.

Abbreviation Index

A

ACS: Automatic Control Systems
ADSL: Asymmetric Digital Subscriber Line
AES: Advanced Encryption Standard
AODV: Ad hoc On–Demand Distance Vector
AOSP: Android Open Source Project
API: Application Programming Interface
APK: Android Package

B

BBU: Base Band Unit
BCH: Broadcast Channel
BER: Bit Error Rate
BF: Beamforming
BLER: Block Error Rate
BS: Base Station

C

CAE: Chinese Academy of Engineering
CBS–HR: Constant Bandwidth Server with Hard–Reservation
CCDF: Complementary Cumulative Distribution Function
CDD: Cyclic Delay Diversity
CDMA: Code Division Multiple Access
CI: Carrier Interferometry
CIR: Channel Impulse Response
CMA: Context Management Architecture
CNGI: China Next Generation Internet
CORAS: Context–Aware Radio Network Simulator
CP: Cyclic Prefix
CQE: Context Quality Enabler
CSCF: Call Session Control Function
CTSE: Cloud–Based Telecommunications Service Environment

D

DAB: Digital Audio Broadcasting
DARPA: Defense Advanced Research Projects Agency
DFT: Discrete Fourier Transform
DHT: Distributed Hash Table
DNS: Domain Name System
DoS: Denial–of–Service
DPD: Digital Pre–Distortion
D–PT: Dynamic Power Tracking
DSL: Digital Subscriber Line
DSP: Digital Signal Processor
DSR: Dynamic Source Routing
DVB: Digital Video Broadcasting

E

ECDH: Elliptic Curve Diffie–Hellman
EDF: Earliest Deadline First
EoC: Ethernet over Coax
EPON: Ethernet Passive Optical Network
ET: Envelope Tracking
ETX: Expected Transmission Count

F

FACH: Forward Access Channel
FDMA: Frequency Division Multiple Access
FDMMSE: Frequency–Domain MMSE
FTTB: Fiber to the Building
FTTZ: Fiber to the Zone

G

GFS: Google File System
GGSN: Gateway GPRS Support Node
GPON: Gigabit Passive Optical Network
GPRS: General Packet Radio Service
GPU: Graphics Processing Unit

H

HAM: Heterogeneous Access Management
HDFS: Hadoop Distributed File System
HPA: High Power Amplifier
HSS: Home Subscriber Server

I

IaaS: Infrastructure as a Service
ICMP: Internet Control Message Protocol
ICT: Information and Communication Technology
IDFT: Inverse Discrete Fourier Transform
IFFT: Inverse Fast Fourier Transform
IRNA: Intelligent Radio Network Access
ISCP: Interferencing Signal Code Power
iSCSI: Internet Small Computer System Interface
ISI: Intersymbol Interference
ISP: Internet Service Provider
IVR: Interactive Voice Response

J

JIT: Just in Time

L

LAN: Local Area Network
LDMOS: Laterally Diffused Metal Oxide Semiconductor
LTE: Long Term Evolution

M

MAC: Media Access Control
MAN: Metropolitan Area Network
MANET: Multihop Mobile Ad hoc Network
MCC: Mobile Cloud Computing
MCL: Mesh Connectivity Layer
MicroTCA: Micro Telecommunications Computing Architecture
MIMO: Multiple Input and Multiple Output
MITM: Man–in–the–Middle
MPU: Memory Protection Unit

N

NE: Network Element

NGB: Next Generation Broadcasting
NIC: Network Interface Card
NMS: Network Management System

O

OFDM: Orthogonal Frequency Division Multiplexing
OFDMA: Orthogonal Frequency Division Multiple Access
OS: Operating System
OSIC: Ordered Successive Interference Cancellation
OSN: Online Social Networking

P

PA: Power Amplifier
PaaS: Platform as a Service
PAPR: Peak-to-Average Power Ratio
PCH: Paging Channel
PNIC: Physical Network Interface Card

Q

QoS: Quality of Service

R

RAT: Radio Access Technology
RF: Radio Frequency
RFID: Radio Frequency Identification
ROM: Read-Only Memory
RU: Radio Frequency Unit

S

SaaS: Software as a Service
SAN: Storage Area Network
SD: Spatial Diversity
SDR: Software Defined Radio
SGSN: Serving GPRS Support Node
SM: Spatial Multiplexing
SNR: Signal-to-Noise Ratio
SOA: Service-Oriented Architectures
SOC: Security Operation Center
SoC: System on Chip
SSL: Secure Sockets Layer

T

TDD: Time Division Duplex
TDMA: Time Division Multiple Access
TDMMSE: Time-Domain MMSE
TD-SCDMA: Time Division-Synchronous Code Division Multiple Access
TLS: Transport Layer Security
TPM: Trusted Platform Module
TSP: Telecommunications Service Provider

U

UDP: User Datagram Protocol
UMTS: Universal Mobile Telecommunications System
UWB: Ultra-Wideband

V

VHDSL: Very High bit rate Digital Subscriber Line
VM: Virtual Machine
VMM: Virtual Machine Manager
VNIC: Virtual Network Interface Card
VPN: Virtual Private Network

W

WAN: Wide Area Network
WAP: Wireless Application Protocol
WCDMA: Wideband Code Division Multiple Access
WCETT: Weighted Cumulative Expected Transmission Time
Wi-Fi: Wireless Fidelity
WML: Wireless Markup Language
WPAN: Wireless Personal Area Networks
WSN: Wireless Sensor Network

Z

ZP: Zero-Padding
ZP-CI: Zero-Padding Carrier Interferometry

Roundup

ZTE Showcases Industry's First MAB Live Demo at Mobile World Congress 2011

ZTE Corporation announced on February 17 that it has successfully demonstrated the Industry's first ever Multiple Access Binding Solution (MAB) in a live demo at Mobile World Congress 2011 in Barcelona.

Branded ALL-JOIN, the MAB solution signifies how a UE can bind multiple RATs simultaneously. The live demo was witnessed by more than 35 top operators from around the world. The Uni Core team successfully demonstrated binding of LTE & Wi-Fi on a single UE through unified authentication by ZTE USPP (UDC). It then bound multiple access technologies using Integrated Serving Gateway (ISGW) built on the advanced service router platform.

During the live demo, operators showed great interest whilst observing a streamed video session established over LTE and FTP Download (P2P traffic case) over Wi-Fi, carried out at the same time. The operators witnessed "the super tunnel" creation in the ISGW and significant bandwidth boost was recorded. Leading operators from Europe, Middle East, India & Asia highly appreciated the live demo and remarked on ZTE's core strategy as "leading innovation".

MAB solution is an effective way to solve data storm issues whilst increasing the revenue for operators at lower investment. MAB can help operators provide the highest quality of services by introducing synergy between their access technologies, as well as by supporting access bindings of any combination for 3G, LTE, Wi-Fi, WiMAX, and Femto. (ZTE Corporation)